(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 766 903**

(51) Int. CI.:

G06Q 10/10	(2012.01)	***G06F 16/242***	(2009.01)
G06Q 40/00	(2012.01)	***G06F 16/2458***	(2009.01)
G06Q 40/02	(2012.01)		
G06Q 20/40	(2012.01)		
G06F 16/28	(2009.01)		
G06F 16/35	(2009.01)		
G06F 16/26	(2009.01)		
G06Q 30/00	(2012.01)		
G06F 16/335	(2009.01)		
G06F 16/9535	(2009.01)		

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **13.03.2014 E 14159535 (5)**

(97) Fecha y número de publicación de la concesión europea: **25.12.2019 EP 2778983**

(54) Título: **Agrupamiento de datos**

(30) Prioridad:

15.03.2013 US 201361800887 P
15.08.2013 US 201313968265
15.08.2013 US 201313968213

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
15.06.2020

(73) Titular/es:

PALANTIR TECHNOLOGIES INC. (100.0%)
100 Hamilton Avenue, Suite 300
Palo Alto, CA 94301, US

(72) Inventor/es:

SPRAGUE, MATTHEW;
KROSS, MICHAEL;
BOROCHOFF, ADAM;
MENON, PARVATHY y
HARRIS, MICHAEL

(74) Agente/Representante:

SÁEZ MAESO, Ana

DESCRIPCIÓN

Agrupamiento de datos

Campo de la invención

Esta invención se refiere al agrupamiento de datos.

Antecedentes de la Invención

En una investigación de fraude, un analista puede tener que tomar decisiones sobre la selección de entidades de datos electrónicos dentro de una recopilación electrónica de datos. Dicha recopilación de datos puede incluir una gran cantidad de elementos de datos (también denominados "entidades de datos") que pueden o no estar relacionados entre sí, y que pueden almacenarse en un almacén electrónico de datos o memoria. Por ejemplo, dicha recopilación de datos puede incluir cientos de miles, millones, decenas de millones, cientos de millones, o incluso miles de millones de entidades de datos, y puede consumir una cantidad considerable de almacenamiento y/o memoria. La determinación y selección de entidades de datos relevantes dentro de dicha recopilación de datos puede ser extremadamente difícil para el analista. Además, el procesamiento de una colección de datos tan grande (por ejemplo, cuando un analista usa un ordenador para filtrar y/o buscar a través de un gran número de entidades de datos) puede ser extremadamente ineficiente y consumir recursos significativos de procesamiento y/o memoria.

La referencia está dirigida a los documentos WO 2008/011728A1, US 6 567 936 B1yWiggerts T.A. - Baxter y otros ("Using clustering algorithms in legacy systems remodularization" INGENIERÍA INVERSA, 1997. PROCEDIMIENTOS DE LA CUARTA CONFERENCIA DE TRABAJO EN AMSTERDAM, PAÍSES BAJOS 6 al 8 DE OCTUBRE DE 1997, LOS ALAMITOS, CA, EE. UU., Sociedad de Computación IEEE, EE. UU., 6 de octubre de 1997 (1997-10-06), páginas 33-43)

Resumen de la invención

La invención se expone en las reivindicaciones 1, 6 y 7 adjuntas, definiéndose características opcionales en las reivindicaciones dependientes adjuntas a las mismas.

Las modalidades de la presente descripción se refieren a la generación automática de estructuras de datos agrupados eficientes en memoria y, más específicamente, a la selección automática de una entidad de datos inicial de interés, agregando la entidad de datos inicial a la estructura de datos agrupados eficiente en memoria (que puede ser denominado en la presente descripción como un "grupo"), y determinando y agregando una o más entidades de datos relacionadas al grupo. En diversas modalidades, un grupo generado puede incluir muchas menos entidades de datos que la recopilación de datos descrita anteriormente, y las entidades de datos incluidas en el grupo solo pueden incluir aquellas entidades de datos que son relevantes para una investigación particular (por ejemplo, una investigación de fraude). En consecuencia, en una modalidad, el procesamiento del grupo generado puede ser altamente eficiente en comparación con la recopilación de datos descritos anteriormente. Esto puede deberse a que, por ejemplo, una investigación de fraude dada por un analista (por ejemplo, cuando el analista tamiza y/o busca a través de entidades de datos de un grupo) solo puede requerir el almacenamiento en la memoria de una única estructura de datos de agrupamiento. Además, un número de entidades de datos en un grupo pueden ser varias órdenes de magnitud más pequeñas que en toda la recopilación electrónica de datos descrita anteriormente porque solo las entidades de datos relacionadas entre sí están incluidas en el grupo.

De acuerdo con diversas modalidades, la presente descripción describe métodos y sistemas mediante los cuales se pueden generar estructuras de datos agrupados de memoria eficiente de entidades de datos relacionadas (o "grupos"). La generación de grupos puede comenzar por la generación, determinación y/o selección automática de una entidad de datos inicial de interés, llamada "semilla". Una entidad de datos puede incluir cualquier dato, información o cosas, como una persona, un lugar, una organización, una cuenta, un ordenador, una actividad y un evento, y/o similares. Las semillas pueden seleccionarse/generarse automáticamente de acuerdo con diversas estrategias de determinación de semillas, y los grupos de entidades de datos relacionadas pueden generarse en función de esas semillas y de acuerdo con las estrategias de generación de grupos (o estrategias de agrupamiento). El sistema puede generar además una puntuación, múltiples puntuaciones y/o meta puntuaciones para cada grupo generado, y opcionalmente puede clasificar o priorizar los grupos generados en función de los meta puntuaciones generados. Los grupos de alta prioridad pueden ser de mayor interés para un analista, ya que pueden contener entidades de datos relacionadas que cumplen con criterios particulares relacionados con la investigación del analista. En una modalidad, el sistema puede permitir que un analista inicie ventajosamente una investigación con un grupo priorizado que incluye muchas entidades de datos relacionadas en lugar de una sola entidad de datos seleccionada al azar. Además, como se describió anteriormente, los requisitos de procesamiento de la investigación del analista pueden ser altamente eficientes en comparación con el procesamiento de la gran colección de datos descritos anteriormente. Como se mencionó anteriormente, esto se debe a que, por ejemplo, una investigación dada por un analista solo puede requerir almacenamiento en la memoria de un solo grupo y, además, un número de entidades de datos en un grupo pueden ser varias órdenes de magnitud más pequeñas que en el conjunto recopilación electrónica de datos descrito anteriormente porque solo las entidades de datos relacionadas entre sí se incluyen en el grupo.

Ventajosamente, de acuerdo con diversas modalidades, las técnicas descritas proporcionan un punto de partida más efectivo para una investigación de entidades de datos de diversos tipos. Un analista puede comenzar una investigación desde un grupo de entidades de datos relacionadas en lugar de una entidad de datos individual, lo que puede reducir la cantidad de tiempo y esfuerzo necesarios para realizar la investigación. Las técnicas descritas también pueden, de acuerdo con diversas modalidades, proporcionar una priorización de grupos múltiples. Por ejemplo, el analista también puede comenzar la investigación desde un grupo de alta prioridad, lo que puede permitirle enfocarse en las investigaciones más importantes. En cada caso, los requisitos de procesamiento y memoria de tal investigación pueden reducirse significativamente debido a la creación y uso de estructuras de datos de agrupamiento altamente eficientes de entidades de datos relacionadas.

Breve descripción de los dibujos

De manera que las características mencionadas anteriormente de la presente invención puedan entenderse en detalle, una descripción más particular de la invención, resumida brevemente anteriormente, puede tener como referencia a las modalidades de la misma que se ilustran en los dibujos adjuntos. Sin embargo, debe observarse que los dibujos adjuntos ilustran solo modalidades preferidas típicas de esta invención y, por lo tanto, no deben considerarse limitantes de su alcance, ya que la invención puede admitir otras modalidades igualmente efectivas.

La Figura 1 es un diagrama de bloques que ilustra un ejemplo de sistema de análisis de datos, de acuerdo con una modalidad.

La Figura 2 ilustra la generación de grupos por el sistema de análisis de datos, de acuerdo con una modalidad.

Las Figuras 3A-3C ilustran el crecimiento de un grupo de entidades de datos relacionadas, de acuerdo con una modalidad.

La Figura 4 ilustra la clasificación de los grupos por el sistema de análisis de datos, de acuerdo con una modalidad de la presente invención.

La Figura 5 ilustra un ejemplo de interfaz de usuario (UI) de análisis de agrupamiento, de acuerdo con una modalidad.

La Figura 6 es un diagrama de flujo de las etapas del método para generar grupos, de acuerdo con una modalidad.

La Figura 7 es un diagrama de flujo de las etapas del método para puntuar grupos, de acuerdo con una modalidad.

La Figura 8 ilustra los componentes de un sistema informático de servidor, de acuerdo con una modalidad.

Descripción Detallada

De acuerdo con diversas modalidades, la presente descripción describe métodos y sistemas mediante los cuales se pueden generar estructuras de datos agrupados de memoria eficiente de entidades de datos relacionadas (o "grupos"). La generación de grupos puede comenzar por la generación, determinación y/o selección automática de una entidad de datos inicial de interés, llamada "semilla". Como se mencionó anteriormente, una entidad de datos puede incluir cualquier dato, información o cosas, como una persona, un lugar, una organización, una cuenta, un ordenador, una actividad y un evento, y/o similares. Las semillas pueden seleccionarse/generarse automáticamente de acuerdo con diversas estrategias de determinación de semillas, y los grupos de entidades de datos relacionadas pueden generarse en función de esas semillas y de acuerdo con las estrategias de generación de grupos (o estrategias de agrupamiento). Se puede acceder a las semillas y a las entidades de datos relacionadas desde varias bases de datos y fuentes de datos que incluyen, por ejemplo, bases de datos mantenidas por instituciones financieras, entidades gubernamentales, entidades privadas, entidades públicas y/o fuentes de datos disponibles públicamente. Dichas bases de datos y fuentes de datos pueden incluir una variedad de información y datos, como, por ejemplo, información personal, información financiera, información relacionada con los impuestos, datos relacionados con la red informática y/o datos de actividades relacionadas con el ordenador, entre otros. Además, las bases de datos y las fuentes de datos pueden incluir varias relaciones que vinculan y/o asocian entidades de datos entre sí. Se pueden almacenar varias entidades y relaciones de datos en diferentes sistemas controlados por diferentes entidades y/o instituciones. De acuerdo con diversas modalidades, el sistema de análisis de datos puede reunir datos de múltiples fuentes de datos para construir grupos.

En diversas modalidades, los métodos y sistemas de la presente descripción pueden permitir a un usuario realizar eficientemente análisis e investigaciones de diversas entidades de datos y grupos de entidades de datos. Por ejemplo, el sistema puede permitir que un usuario (también denominado en la presente descripción "analista") realice varias investigaciones financieras y de seguridad relacionadas con una semilla (por ejemplo, una entidad de datos inicial u objeto de datos). Como se describió anteriormente, los requisitos de procesamiento y memoria de dicha investigación pueden reducirse significativamente debido a la creación y uso de estructuras de datos de agrupamiento altamente eficientes de entidades de datos relacionadas. Para realizar investigaciones financieras y de seguridad relacionadas con la semilla, un analista puede tener que buscar varias capas de entidades de datos relacionadas. Por ejemplo, el analista podría investigar entidades de datos relacionadas con una cuenta de tarjeta de crédito inicial, descubriendo los identificadores de cliente asociados con la cuenta de tarjeta de crédito, los números de teléfono asociados con esos

identificadores de clientes, los identificadores de clientes adicionales asociados con esos números de teléfono y finalmente las cuentas de tarjetas de crédito adicionales asociadas con esos identificadores de clientes adicionales. Si la cuenta inicial de la tarjeta de crédito fuera fraudulenta, entonces el analista podría determinar que las cuentas adicionales de la tarjeta de crédito también podrían ser fraudulentas. En dicha investigación, el analista descubriría la relación entre las cuentas de tarjetas de crédito adicionales y las cuentas de tarjetas de crédito semilla a través de varias capas de entidades de datos relacionadas. Esta técnica es particularmente valiosa para investigaciones en las que la relación entre entidades de datos podría incluir varias capas y sería difícil de identificar.

En una modalidad, el sistema de análisis de datos descubre automáticamente entidades de datos relacionadas con una semilla y almacena las relaciones resultantes y las entidades de datos relacionadas juntas en un "grupo". Una estrategia de generación de grupo especifica qué búsquedas realizar en cada etapa del proceso de investigación. Las búsquedas producen capas de entidades de datos relacionadas para agregar al grupo. Por lo tanto, el analista comienza una investigación con el grupo resultante, en lugar de la semilla sola. Comenzando con el grupo, el analista puede formarse opiniones con respecto a las entidades de datos relacionadas, realizar un análisis adicional de las entidades de datos relacionadas o puede consultar entidades de datos relacionadas adicionales. Además, para numerosas semillas de este tipo e investigaciones asociadas, el sistema de análisis de datos puede priorizar los grupos basados en una agregación de características de las entidades de datos relacionadas dentro de los grupos. El sistema de análisis de datos luego muestra resúmenes de los grupos. Los resúmenes pueden mostrarse de acuerdo con la priorización. La priorización puede ayudar al analista a seleccionar qué grupos investigar.

En la siguiente descripción, se exponen numerosos detalles específicos con el objetivo de proporcionar una comprensión total de la presente invención. Sin embargo, será evidente para un experto en la técnica que la presente invención puede llevarse a la práctica sin uno a más de estos detalles específicos.

La Figura 1 es un diagrama de bloques que ilustra un ejemplo de sistema de análisis de datos 100, de acuerdo con una modalidad de la presente invención. Como se muestra, el sistema de análisis de datos 100 incluye un servidor de aplicaciones 115 que se ejecuta en un sistema informático de servidor 110, un cliente 135 que se ejecuta en un sistema informático de cliente 130 y al menos una base de datos 140. Además, el cliente 135, el servidor de aplicaciones 115 y la base de datos 140 pueden comunicarse a través de una red 150, por ejemplo, para acceder a las fuentes de datos del grupo 160.

El servidor de aplicaciones 115 incluye un motor de agrupamiento 120 y un motor de flujo de trabajo 125. El motor de agrupamiento 120 se configura para construir uno o más grupos de entidades de datos relacionadas, de acuerdo con una estrategia de análisis definida. El motor de agrupamiento 120 puede leer datos de una variedad de fuentes de datos de agrupamiento 160 para generar grupos a partir de datos semilla. Una vez creados, los grupos resultantes pueden almacenarse en el ordenador servidor 110 o en la base de datos 140. Las operaciones del motor de agrupamiento 120 se analizan en detalle a continuación junto con las Figuras 2 y 3.

El motor de agrupamiento 120 se configura para puntuar los grupos, de acuerdo con una estrategia de puntuación definida. La puntuación puede indicar la importancia de analizar el grupo. Por ejemplo, el motor de agrupamiento 120 podría ejecutar una estrategia de puntuación que agregue los saldos de cuentas de tarjetas de crédito dentro del grupo. Si el grupo incluye un saldo total mayor que otros grupos, entonces el grupo podría ser una responsabilidad mayor para la institución financiera. Por lo tanto, el grupo sería más importante para analizar y recibiría una puntuación más alta. En una modalidad, el motor de agrupamiento 120 organiza y presenta los grupos de acuerdo con las puntuaciones asignadas. El motor de agrupamiento 120 puede presentar resúmenes de los grupos y/o representaciones interactivas de los grupos dentro de la IU de análisis de agrupamiento. Por ejemplo, las representaciones pueden proporcionar gráficos visuales de las entidades de datos relacionadas dentro de los grupos. El motor de agrupamiento 120 puede generar la IU de análisis de agrupamiento como una aplicación web o una página web dinámica mostrada dentro del cliente 135. El motor de agrupamiento 120 también permite que un analista cree tareas asociadas con los grupos. Las operaciones del motor de agrupamiento 120 se analizan en detalle a continuación junto con las Figuras 4 y 5. En una modalidad, el motor de agrupamiento 120 genera grupos automáticamente, para su posterior revisión por parte de los analistas. Los analistas también pueden asignarse tareas a sí mismos a través de una IU de flujo de trabajo. El motor de flujo de trabajo 125 consume puntuaciones generadas por el motor de agrupamiento 120. Por ejemplo, el motor de flujo de trabajo 125 puede presentar un analista con grupos generados, puntuados y ordenados por el motor de agrupamiento 120.

El cliente 135 representa una o más aplicaciones informáticas configuradas para presentar datos y traducir la entrada, desde el analista, en solicitudes de análisis de datos por parte del servidor de aplicaciones 115. En una modalidad, el cliente 135 y el servidor de aplicaciones 115 están acoplados entre sí. Sin embargo, varios clientes 135 pueden ejecutarse en el ordenador cliente 130 o varios clientes 135 en varios ordenadores cliente 130 pueden interactuar con el servidor de aplicaciones 115. En una modalidad, el cliente 135 puede ser un navegador que accede a un servicio web.

Mientras que el cliente 135 y el servidor de aplicaciones 115 se muestran ejecutándose en distintos sistemas informáticos, el cliente 135 y el servidor de aplicaciones 115 pueden ejecutarse en el mismo sistema informático. Además, el motor de agrupamiento 120 y el motor de flujo de trabajo 125 pueden ejecutarse en servidores de aplicaciones separados 115, en sistemas informáticos de servidor separados, o alguna combinación de los mismos. Adicionalmente, un servicio de historial puede almacenar los resultados generados por un analista en relación con un grupo determinado

En una modalidad, las fuentes de datos de agrupamiento 160 proporcionan datos disponibles para el motor de agrupamiento para crear grupos a partir de un conjunto de semillas. Dichas fuentes de datos pueden incluir fuentes de datos relacionales, datos de servicios web, datos XML, etc. Por ejemplo, las fuentes de datos pueden relacionarse con registros de cuentas de clientes almacenados por una institución financiera. En tal caso, las fuentes de datos pueden incluir datos de cuenta de tarjeta de crédito, datos de cuenta bancaria, datos de clientes y datos de transacciones. Los datos pueden incluir atributos de datos tales como números de cuenta, saldos de cuenta, números de teléfono, direcciones y montos de transacciones, etc. Por supuesto, las fuentes de datos de agrupamiento 160 se incluyen para ser representativas de una variedad de datos disponibles para el sistema informático del servidor 110 a través de la red 150, así como las fuentes de datos disponibles localmente.

La base de datos 140 puede ser un Sistema de gestión de bases de datos relacionales (RDBMS) que almacena los datos como filas en tablas relacionales. Mientras que la base de datos 140 se muestra como un sistema informático distinto, la base de datos 140 puede funcionar en el mismo sistema informático del servidor 110 que el servidor de aplicaciones 115.

La Figura 2 ilustra la generación de grupos por el sistema de análisis de datos 200, de acuerdo con una modalidad. Como se muestra, el sistema de análisis de datos 200 interactúa con una lista de semillas 210, una lista de grupos 250 y un almacén de estrategias de agrupamientos 230. La lista de semillas 210 incluye semillas 212-1, 212-2 ... 212-S y la lista de grupos 250 incluye los grupos 252-1, 252-2 ... 252-C. El motor de agrupamiento 120 se configura como una aplicación informática o hilo que genera los grupos 252-1, 252-2 ... 252-C de las semillas 212-1, 212-2 ... 212-S.

Las semillas 212 son el punto de partida para generar un grupo 252. Para generar un grupo, el motor de agrupamiento 120 recupera una semilla 212 dada de la lista 210 de semillas. La semilla 212 puede ser una entidad de datos arbitraria dentro de la base de datos 140, tal como un nombre de cliente, un número de seguro social del cliente, un número de cuenta o un número de teléfono del cliente.

El motor de agrupamiento 120 genera el grupo 252 a partir de la semilla 212. En una modalidad, el motor de agrupamiento 120 genera el grupo 252 como una colección de entidades de datos y las relaciones entre las diversas entidades de datos. Como se señaló anteriormente, la estrategia de agrupamiento ejecuta enlaces de datos para agregar cada capa adicional de objetos al grupo. Por ejemplo, el motor de agrupamiento 120 podría generar el grupo 252 a partir de una cuenta de tarjeta de crédito inicial. El motor de agrupamiento 120 primero agrega la cuenta de tarjeta de crédito al grupo 252. El motor de agrupamiento 120 podría agregar clientes relacionados con la cuenta de tarjeta de crédito al grupo 252. El motor de agrupamiento 120 podría completar el grupo 252 agregando cuentas de tarjetas de crédito adicionales relacionadas con esos clientes. A medida que el motor de agrupamiento 120 genera el grupo 252, el motor de agrupamiento 120 almacena el grupo 252 dentro de la lista de grupos 250. El grupo 252 puede almacenarse como una estructura de datos de gráfico. La lista de agrupamiento 250 puede ser una colección de tablas en la base de datos 140. En tal caso, puede haber una tabla para las entidades de datos del grupo 252, una tabla para las relaciones entre las diversas entidades de datos, una tabla para los atributos de las entidades de datos y una tabla para una puntuación del grupo 252 . La lista de grupos 250 puede incluir grupos 252 de múltiples investigaciones. Tenga en cuenta que el motor de agrupamiento 120 puede almacenar partes del grupo 252 en la lista de grupos 250 a medida que el motor de agrupamiento 120 genera el grupo 252.

Los expertos en la técnica reconocerán que existen muchas técnicas técnicamente viables para crear y almacenar estructuras de datos gráficos.

El depósito de estrategias de agrupamiento 230 incluye estrategias de agrupamiento 232-1, 232-2 ... 232-N. Cada estrategia de agrupamiento puede incluir referencias 235 a uno o más enlaces de datos 237. Como se señaló, cada enlace de datos puede usarse para identificar datos que pueden hacer crecer un grupo (según lo determinado por la estrategia de búsqueda dada 232). El motor de agrupamiento 120 ejecuta una estrategia de agrupamiento 232 para generar el grupo 252. Específicamente, el motor de agrupamiento 120 ejecuta la estrategia de agrupamiento 232 seleccionada por un analista. El analista puede enviar una selección de la estrategia de agrupamiento 232 al motor de agrupamiento 120 a través del cliente 135.

Cada estrategia de agrupamiento 232 se configura para realizar procesos de investigación para generar el grupo 252. Nuevamente, por ejemplo, la estrategia de agrupamiento 232 puede incluir referencias 235 a una colección de enlaces de datos ejecutados para agregar capa tras capa de datos a un grupo. El proceso de investigación incluye búsquedas para recuperar entidades de datos relacionadas con la semilla 212. Por ejemplo, la estrategia de agrupamiento 232 podría comenzar con una cuenta de tarjeta de crédito posiblemente fraudulenta como la semilla 212. La estrategia de agrupamiento 232 buscaría clientes relacionados con la cuenta de tarjeta de crédito, y luego cuentas de tarjeta de crédito adicionales relacionadas con esos clientes. Una estrategia de agrupamiento diferente 232 podría buscar clientes relacionados con la cuenta de la tarjeta de crédito, números de teléfono relacionados con los clientes, clientes adicionales relacionados con los números de teléfono y cuentas de tarjetas de crédito adicionales relacionadas con los clientes adicionales.

En una modalidad, la estrategia de agrupamiento 232 incluye una referencia a al menos un enlace de datos 237. El motor de agrupamiento 120 ejecuta el protocolo de búsqueda especificado por el enlace de datos 237 para recuperar datos, y

los datos devueltos por un enlace de datos dado forman una capa dentro del grupo 252. Por ejemplo, el enlace de datos 237 podría recuperar conjuntos de clientes relacionados con una cuenta por un atributo del propietario de la cuenta. El enlace de datos 237 recupera el conjunto de entidades de datos relacionadas de una fuente de datos. Por ejemplo, el enlace de datos 237-1 podría definir la especificación de una consulta de base de datos para realizar contra una base de datos. Del mismo modo, el enlace de datos 237-2 podría definir una conexión a un sistema remoto de base de datos relacional y el enlace de datos 237-3 podría definir una conexión y consulta contra un servicio web de terceros. Una vez recuperada, la estrategia de agrupamiento 232 puede evaluar si los datos devueltos deberían agregarse a un grupo que se cultiva a partir de una semilla 212 dada. Múltiples estrategias de agrupamiento 232 pueden hacer referencia a un enlace de datos dado 237. El analista puede actualizar el enlace de datos 237, pero generalmente actualiza el enlace de datos 237 solo si cambia la fuente de datos asociada. Una estrategia de agrupamiento 232 también puede incluir un enlace de datos dado 237 múltiples veces. Por ejemplo, ejecutar un enlace de datos 237 mediante el uso de una semilla 212 puede generar semillas adicionales para ese enlace de datos 237 (o generar semillas para otro enlace de datos 237). Más generalmente, diferentes estrategias de agrupamiento 232-1, 232-2 ... 232-N puede incluir diferentes disposiciones de varios enlaces de datos 237 para generar diferentes tipos de grupos 252.

La estrategia de agrupamiento 232 puede especificar que el motor de agrupamiento 120 use un atributo de las entidades de datos relacionadas recuperadas con un enlace de datos 237, como entrada a un enlace de datos posterior 237. El motor de agrupamiento 120 utiliza el enlace de datos posterior 237 para recuperar una capa posterior de entidades de fecha relacionadas para el grupo 252. Por ejemplo, la estrategia de agrupamiento 232 podría especificar que el motor de agrupamiento 120 recupere un conjunto de entidades de datos de cuenta de tarjeta de crédito con un primer enlace de datos 237-1. La estrategia de agrupamiento 232 también podría especificar que el motor de agrupamiento 120 luego utilice el atributo de número de cuenta de entidades de datos de cuenta de tarjeta de crédito como entrada para un enlace de datos posterior 237-2. La estrategia de agrupamiento 232 también puede especificar filtros para que el motor de agrupamiento 120 se aplique a los atributos antes de realizar el enlace de datos posterior 237. Por ejemplo, si el primer enlace de datos 237-1 fuera recuperar un conjunto de entidades de datos de cuentas de tarjetas de crédito que incluían cuentas de tarjetas de crédito personales y comerciales, entonces el motor de agrupamiento 120 podría filtrar las cuentas de tarjetas de crédito comerciales antes de realizar los datos posteriores vinculante 237-2.

En funcionamiento, el motor de agrupamiento 120 genera un grupo 252-1 a partir de una semilla 212-1 recuperando primero una estrategia de agrupamiento 232. Suponiendo que el analista seleccionó una estrategia de agrupamiento 232-2, entonces el motor de agrupamiento 120 recuperaría la estrategia de agrupamiento 232-2 del almacén de estrategia de agrupamiento 230. El motor de agrupamiento 120 podría recuperar la semilla 212-1 como entrada a la estrategia de agrupamiento 232-2. El motor de agrupamiento 120 ejecutaría la estrategia de agrupamiento 232-2 recuperando conjuntos de datos ejecutando enlaces de datos 237 referenciados por la estrategia de agrupamiento 232-2. Por ejemplo, la estrategia de agrupamiento podría ejecutar enlaces de datos 237-1, 237-2 y 237-3. El motor de agrupamiento 120 evalúa los datos devueltos por cada enlace de datos 237 para determinar si usar esos datos para hacer crecer el grupo 252-1. El motor de agrupamiento 120 puede usar elementos de los datos devueltos como entrada para el siguiente enlace de datos 237. Por supuesto, una variedad de rutas de ejecución son posibles para los enlaces de datos 237. Por ejemplo, suponga que un enlace de datos 237 devolvió un conjunto de números de teléfono. En tal caso, otro enlace de datos 237 podría evaluar cada número de teléfono individualmente. Como otro ejemplo, un enlace de datos 237 podría usar parámetros de entrada obtenidos ejecutando enlaces de datos múltiples y otros 237. De manera más general, el motor de agrupamiento 120 puede recuperar datos para cada enlace de datos referenciado por la estrategia de agrupamiento 232-2. El motor de agrupamiento 120 almacena el grupo completo 252-1 en la lista de grupos 250.

A medida que el motor de agrupamiento 120 genera los grupos 252-1, 252-2 ... 252-C de semillas 212-1, 212-2 ... 212-S, la lista de grupos 250 puede incluir grupos superpuestos 252. Dos grupos 252-1 y 252-C se superponen si ambos grupos 252-1 y 252-C incluyen una entidad de datos común. A menudo, un grupo más grande 252 formado al fusionar dos grupos más pequeños 252-1 y 252-C puede ser un mejor punto de partida de investigación que los grupos más pequeños 252-1 y 252-C individualmente. El grupo más grande 252 puede proporcionar información o relaciones adicionales, que pueden no estar disponibles si los dos grupos 252-1 y 252-C permanecen separados.

En una modalidad, el motor de agrupamiento 120 incluye un solucionador 226 que se configura para detectar y fusionar dos grupos superpuestos 252 juntos. El solucionador 226 compara las entidades de datos dentro de un grupo 252-1 con las entidades de datos dentro de cada uno de los otros grupos 252-2 a 252-C. Si el solucionador 226 encuentra la misma entidad de datos dentro del grupo 252-1 y un segundo grupo 252-C, entonces el solucionador 226 puede fusionar los dos grupos 252-1 y 252-C en un único grupo más grande 252. Por ejemplo, el grupo 252-1 y el grupo 252-C podrían incluir al mismo cliente. El solucionador 226 compararía las entidades de datos del grupo 252-1 con las entidades de datos del grupo 252-C y detectaría el mismo cliente en ambos grupos 252. Al detectar el mismo cliente en ambos grupos 252, el solucionador 226 podría fusionar el grupo 252-1 con el grupo 252-C. El solucionador 226 puede probar cada par de grupos 252 para identificar grupos superpuestos 252. Aunque los grupos más grandes 252 pueden ser mejores puntos de partida de investigación, un analista puede querer entender cómo el solucionador 226 formó los grupos más grandes 252. El solucionador 226 almacena un historial de cada fusión.

Después de que el motor de agrupamiento genera un conjunto de grupos a partir de una colección dada de semillas (y después de fusionar o resolver el grupo), el motor de agrupamiento 120 puede puntuar, clasificar u ordenar los grupos en relación con una estrategia de puntuación 442.

En una modalidad, el sistema de análisis 100, y más específicamente, el motor de agrupamiento 120 recibe una lista de semillas para generar un conjunto de grupos, posteriormente clasificadas, ordenadas y presentadas a analistas. Es decir, el motor de agrupamiento 120 consume semillas generadas por otros sistemas. Alternativamente, en otras modalidades, el motor de agrupamiento 120 puede generar las semillas 212-1, 212-2 ... 212-S. Por ejemplo, el motor de agrupamiento 120 puede incluir una estrategia de generación principal que identifica entidades de datos como posibles semillas 212. La estrategia de generación principal puede aplicarse a un tipo de negocio en particular, como tarjetas de crédito, negociación de acciones o reclamos de seguros, y puede ejecutarse contra una fuente de datos del grupo 160 o una fuente de información externa.

Las Figuras 3A - 3C ilustran el crecimiento de un grupo 252 de entidades de datos relacionadas, de acuerdo con una modalidad. Como se muestra en la Figura 3A, un grupo 252 incluye una entidad de datos semilla 302, enlaces 303-1 y 303-2, y entidades de datos relacionadas 305-1 y 305-2. El grupo 252 se basa en una semilla 212. El motor de agrupamiento 120 construye el grupo 252 ejecutando una estrategia de agrupamiento 232 con las siguientes búsquedas:

- Encuentra propietario de semilla
- Encuentra todos los números de teléfono relacionados con el propietario de la semilla
- Encuentra todos los clientes relacionados con los números de teléfono
- Encuentra todas las cuentas relacionadas con los clientes
- Encuentra todos los nuevos clientes relacionados con las nuevas cuentas

Suponiendo que la semilla 212 fuera una cuenta de tarjeta de crédito fraudulenta, entonces el motor de agrupamiento 120 agregaría la cuenta de tarjeta de crédito al grupo 252 como la entidad de datos semilla 302. El motor de agrupamiento 120 usaría entonces el atributo del propietario de la cuenta de la tarjeta de crédito como entrada para un enlace de datos 237. El motor de agrupamiento 120 ejecutaría el protocolo de búsqueda de enlace de datos 237 para recuperar los datos del cliente que identifican al propietario de la cuenta fraudulenta de la tarjeta de crédito. El motor de agrupamiento 120 agregaría entonces los datos del cliente al grupo 252 como la entidad de datos relacionada 305-1. El motor de agrupamiento 120 también agregaría el atributo del propietario de la cuenta como el enlace 303-1 que relaciona el número de cuenta con los datos del cliente del propietario. El motor de agrupamiento 120 ejecutaría la próxima búsqueda de la estrategia de agrupamiento 232 introduciendo el atributo identificador del cliente de los datos del cliente en un enlace de datos 237 para recuperar los datos de un teléfono. El motor de agrupamiento 120 agregaría entonces los datos del teléfono como la entidad de datos relacionada 305-2 y el atributo identificador del cliente como el enlace 303-2 entre los datos del cliente y los datos del teléfono. En este punto del proceso de investigación, el grupo 252 incluiría la entidad de datos semilla 302, dos enlaces 303-1 y 303-2, y dos entidades de datos relacionadas 305-1 y 305-2. Es decir, el grupo 252 incluye la cuenta de tarjeta de crédito fraudulenta, los datos del cliente del propietario de la tarjeta de crédito y el número de teléfono del propietario. Al llevar más lejos el proceso de investigación, el motor de agrupamiento 120 podría revelar más información relacionada, por ejemplo, clientes adicionales o cuentas de tarjetas de crédito potencialmente fraudulentas.

Volviendo a la Figura 3B, el motor de agrupamiento 120 continuaría ejecutando la estrategia de agrupamiento 232 buscando entidades de datos de cuenta adicionales relacionadas con el número de teléfono del propietario de la cuenta de tarjeta de crédito fraudulenta. Como se discutió, el número de teléfono se almacenaría como entidad de datos relacionada 305-2. El motor de agrupamiento 120 ingresaría el atributo del propietario del teléfono del número de teléfono a un enlace de datos 237. El motor de agrupamiento 120 ejecutaría el protocolo de búsqueda de enlace de datos 237 para recuperar los datos de dos clientes adicionales, que el motor de agrupamiento 120 almacenaría como entidades de datos relacionadas 305-3 y 305-4. El motor de agrupamiento 120 agregaría el atributo del propietario del teléfono como los enlaces 303-3 y 304-4 entre los clientes adicionales y el número de teléfono.

La Figura 3C muestra el grupo 252 después de que el motor de agrupamiento 120 realiza la última etapa de la estrategia 232 del grupo. Por ejemplo, el motor de agrupamiento 120 usaría el atributo de identificador de cliente de la entidad de datos relacionada 305-3 y 305-4 para recuperar y agregar entidades de datos de cuenta adicionales como las entidades de datos relacionadas 305-5 y 305-6. El motor de agrupamiento 120 acoplaría las entidades de datos relacionadas 305-5 y 305-6 a las entidades de datos relacionadas 305-3 y 305-4 con los atributos de identificación del cliente almacenados como enlaces 303-5 y 303-6. Por lo tanto, el grupo 252 incluiría seis entidades de datos relacionadas 305 relacionadas por seis enlaces 303, además de la entidad de datos semilla 302. El analista podría identificar y determinar si las entidades de cuentas de datos adicionales, almacenadas como entidades de datos relacionadas 305-3 y 305-4, representan cuentas de tarjetas de crédito fraudulentas de manera más eficiente, que si el analista iniciara una investigación con solo la semilla 212. Como ilustra lo anterior, con el motor de agrupamiento 120 y la estrategia de agrupamiento 232, el analista puede iniciar ventajosamente una investigación desde un grupo 252 que ya incluye varias entidades de datos relacionadas 305.

La Figura 4 ilustra la clasificación de los grupos 252 por el sistema de análisis de datos 100 mostrado en la Figura 1, de acuerdo con una modalidad de la presente invención. Como se muestra, la Figura 4 ilustra algunos de los mismos elementos que se muestran en la Figura 1 y la Figura 2. Además, la Figura 4 ilustra un almacén de estrategia de puntuación 440, acoplado al motor de flujo de trabajo 125.

El motor de agrupamiento 120 acoplado a la lista de grupos 250. El almacén de estrategias de puntuación 440 incluye estrategias de puntuación 442-1, 442-2 ... 442-R.

El motor de agrupamiento 120 ejecuta una estrategia de puntuación 442 para puntuar un grupo 252. Por ejemplo, el motor de agrupamiento 120 puede generar un grupo, a través de una estrategia de agrupamiento/enlaces de datos, e intentar resolverlo con grupos existentes. A partir de entonces, el motor de agrupamiento 120 puede puntuar el agrupamiento resultante con cualquier estrategia de puntuación asociada con una estrategia de generación de agrupamiento dada. En una modalidad, la puntuación para un grupo puede ser una meta puntuación generada como una agregación de puntuaciones generadas para diferentes aspectos, métricas o datos de un grupo. El pedido de un grupo de grupos (de acuerdo con una estrategia de puntuación dada) se puede realizar a pedido cuando un cliente lo solicite. Alternativamente, el analista puede seleccionar una estrategia de puntuación 442 para el motor de agrupamiento 120 a través del cliente 35 o el analista puede incluir la selección dentro de una secuencia de comandos o archivo de configuración. En otras modalidades, el motor de agrupamiento 120 puede ejecutar varias estrategias de puntuación 442 para determinar una puntuación combinada para el grupo 252.

La estrategia de puntuación 442 especifica un enfoque para puntuar un grupo 252. La puntuación puede indicar la importancia relativa de un grupo dado 252. Por ejemplo, el motor de agrupamiento 120 podría ejecutar una estrategia de puntuación 442-1 para determinar una puntuación contando el número de un tipo de entidad de datos particular dentro del grupo 252. Supongamos, por ejemplo, que una entidad de datos corresponde a una cuenta de crédito. En tal caso, un grupo con una gran cantidad de cuentas abiertas por una sola persona (posiblemente en poco tiempo) podría correlacionarse con un mayor riesgo de fraude. Por supuesto, una puntuación de grupo puede estar relacionado con un alto riesgo de fraude basado en los otros datos en el grupo, según sea apropiado para un caso dado. De manera más general, cada estrategia de puntuación 442 puede adaptarse según los datos en grupos creados por una estrategia de agrupamiento 230 dada y el tipo particular de riesgo o fraude (o cantidades en riesgo).

En funcionamiento, el motor de agrupamiento 120 puntúa un grupo 252-1 recuperando primero una estrategia de puntuación 442. Por ejemplo, suponga que un analista selecciona la estrategia de puntuación 442-1. En respuesta, el motor de agrupamiento 120 recupera la estrategia de puntuación 442-1. El motor de agrupamiento 120 también recupera el grupo 252-1 de la lista de grupos 250. Después de determinar la puntuación del grupo 252-1, el motor del grupo 120 puede almacenar la puntuación con el grupo 252-1 en la lista de grupos 250.

El motor de agrupamiento 120 puede puntuar múltiples grupos 252-1, 252-2 ... 252-C en la lista de grupos 250. El motor de agrupamiento 120 también puede clasificar los grupos 252-1, 252-2 ... 252-C basado en los puntajes. Por ejemplo, el motor de agrupamiento 120 podría clasificar el grupo 252-1, 252-2 ... 252-C de la puntuación más alta a la puntuación más baja.

La Figura 5 ilustra un ejemplo de la IU del análisis de agrupamiento 500, de acuerdo con una modalidad. Como se describió, el motor de flujo de trabajo 125 se configura para presentar la IU del análisis de agrupamiento 500. Como se muestra, la IU del análisis de agrupamientos 500 incluye un cuadro principal 510, un cuadro de estrategia de agrupamiento 530, una lista de resumen del grupo 525, un cuadro de búsqueda del grupo 520 y una ventana de revisión de grupos 515. El motor de flujo de trabajo 125 puede generar la IU del análisis de agrupamiento 500 como una aplicación web o una página web dinámica mostrada dentro del cliente 135.

El cuadro principal 510 permite al analista seleccionar una lista de semillas 210 o una estrategia adecuada de generación principal. La estrategia de generación principal genera una lista de semillas 210. La estrategia de generación principal puede generar una lista de semillas 210 de la base de datos 140 o una fuente de información externa (por ejemplo, una fuente de datos del grupo 160).

El cuadro de estrategia de agrupamiento 530 muestra las estrategias de agrupamiento 232 que el motor de agrupamiento 120 ejecutó contra la lista de semillas 210. El motor de agrupamiento 120 puede ejecutar múltiples estrategias de agrupamiento 232 contra la lista de semillas 210, por lo que puede haber múltiples estrategias de agrupamiento 232 enumeradas en el cuadro de estrategia de agrupamiento 530. El analista puede hacer clic en el nombre de una estrategia de agrupamiento dada 232 en el cuadro de estrategia de agrupamiento 530 para revisar los grupos 252 que generó la estrategia de agrupamiento 232.

El motor de flujo de trabajo 125 muestra resúmenes de los grupos 252 en la lista de resumen del grupo 525. Por ejemplo, los resúmenes pueden incluir características de los grupos 252, como identificadores, puntuaciones o analistas asignados para analizar los grupos 252. El motor de flujo de trabajo 125 puede seleccionar los grupos 252 para la visualización en la lista de resumen del grupo 525 de acuerdo con esas u otras características. Por ejemplo, el motor de flujo de trabajo 125 podría mostrar los resúmenes en el orden de las puntuaciones de los grupos 252, donde se muestra primero un resumen del grupo de mayor puntuación 252.

El motor de flujo de trabajo 125 controla el orden y la selección de los resúmenes dentro de la lista de resumen del grupo 525 basándose en la entrada del analista. El cuadro de búsqueda de grupo 520 incluye un cuadro de texto de búsqueda junto con un botón de búsqueda y un control desplegable. El analista puede ingresar una característica de un grupo 252 en el cuadro de texto de búsqueda y luego indicar al motor de flujo de trabajo 125 que busque y muestre grupos 252 que

incluyen la característica presionando el botón de búsqueda. Por ejemplo, el analista podría buscar grupos con una puntuación particular. El control desplegable incluye una lista de diferentes características de los grupos 252, como puntuación, tamaño, analista asignado o fecha de creación. El analista puede seleccionar una de las características para indicar al motor de flujo de trabajo 125 que presente los resúmenes de los grupos 252 dispuestos por esa característica.

El motor de flujo de trabajo 125 también se configura para presentar detalles de un grupo 252 dado dentro de la ventana de revisión de grupos 515. El motor de flujo de trabajo 125 muestra los detalles del grupo 252, por ejemplo, la puntuación o saldos de cuenta promedio dentro de un grupo, cuando el analista hace clic en un puntero del mouse en el resumen asociado dentro de la lista de resumen del grupo 525. El motor de flujo de trabajo 125 puede presentar detalles del grupo 252, tales como el nombre del analista asignado para analizar el grupo 252, la puntuación del grupo 252 y estadísticas o gráficos generados a partir del grupo 252. Estos detalles le permiten al analista determinar si debe investigar más el grupo 252. La ventana de revisión de grupos 515 también incluye un botón en el que se puede hacer clic para investigar un grupo 252 dentro de un gráfico y un botón de asignación para asignar un grupo a un analista.

El analista puede hacer clic en el puntero del ratón sobre el botón para investigar el grupo 252 dentro de un gráfico interactivo. La representación interactiva es un gráfico visual del grupo 252, donde los iconos representan las entidades del grupo 252 y las líneas entre los iconos representan los enlaces entre las entidades del grupo 252. Por ejemplo, el motor de flujo de trabajo 125 podría mostrar el gráfico interactivo del grupo 252 similar a la representación del grupo 252 en la Figura 3C. La representación interactiva permite al analista revisar los atributos de las entidades de datos relacionadas o realizar consultas para entidades de datos relacionadas adicionales.

Un usuario administrativo puede hacer clic en el puntero del ratón en el botón de asignación para asignar el grupo asociado 252 a un analista. El motor de flujo de trabajo 125 también permite al usuario administrativo crear tareas asociadas con los grupos 252, mientras que el usuario administrativo asigna el grupo 252. Por ejemplo, el usuario administrativo podría crear una tarea para buscar en los tres grupos 252 de mayor puntuación cuentas de tarjetas de crédito fraudulentas. El motor de flujo de trabajo 125 puede mostrar los resúmenes en la lista de resumen del grupo 525 de acuerdo con los nombres de los analistas asignados a los grupos 252. Del mismo modo, el motor de flujo de trabajo 125 solo puede mostrar resúmenes para el subconjunto de los grupos 252 asignados a un analista.

La interfaz que se muestra en la Figura 5 se incluye para ilustrar una interfaz ilustrativa útil para navegar y revisar grupos generados mediante el uso del motor de agrupamiento 120 y el motor de flujo de trabajo 125. Por supuesto, un experto en la técnica reconocerá que podría usarse una amplia variedad de construcciones de interfaz de usuario para permitir al analista seleccionar estrategias de agrupamiento 232, estrategias de puntuación 242 o estrategias de generación de semillas, iniciar una investigación o revisar y analizar los grupos 252. Por ejemplo, el motor de flujo de trabajo 125 puede mostrar controles adicionales dentro de la IU del análisis de agrupamiento 500 para controlar el proceso de generación de grupos y seleccionar estrategias de agrupamiento 232 o estrategias de puntuación 242. Además, el motor de flujo de trabajo 125 puede no mostrar el cuadro principal 510 o las opciones para seleccionar una estrategia de generación principal. Además, aunque el motor de flujo de trabajo 125 genera la IU del análisis de agrupamiento 500, en diferentes modalidades, la IU del análisis de agrupamiento 500 es generada por una aplicación informática distinta del motor de flujo de trabajo 125. Además, en diferentes modalidades, la ventana de revisión de grupos 515 está configurada para mostrar una vista previa del grupo 252 o estadísticas adicionales generadas a partir del grupo 252. Como tal, se puede presentar una representación interactiva del grupo 252 en una IU adicional o se puede exportar el grupo 252 a otra aplicación informática para su revisión por el analista.

La Figura 6 es un diagrama de flujo de las etapas del método para generar grupos, de acuerdo con una modalidad. Aunque las etapas del método se describen junto con los sistemas de las Figuras 1 y 2, los expertos en la técnica entenderán que cualquier sistema configurado para realizar las etapas del método, en cualquier orden, está dentro del alcance de la presente invención. Además, el método 600 se puede realizar junto con el método 700 para puntuar un grupo, que se describe a continuación.

Como se muestra, el método 600 comienza en la etapa 605, donde el motor de agrupamiento 120 recupera una estrategia de agrupamiento 232 y una semilla 212. Una vez que se selecciona una estrategia de agrupamiento el motor de agrupamiento 120 identifica una lista de semillas para construir grupos mediante el uso de la estrategia de agrupamiento seleccionada. En la etapa 610, el motor de agrupamiento 120 inicializa un grupo 252 con una de las semillas en la lista. El grupo 252 se almacena como una estructura de datos gráficos. El motor de agrupamiento 120 inicializa la estructura de datos del gráfico, y luego agrega la semilla 212-1 a la estructura de datos del gráfico como la primera entidad de datos.

En la etapa 615, el motor de agrupamiento 120 hace crecer el grupo 252 ejecutando el protocolo de búsqueda de un enlace de datos 237 a partir de la estrategia de agrupamiento 232-2. La estrategia de agrupamiento 232-2 incluye una serie de enlaces de datos 237 que el motor de agrupamiento 120 ejecuta para recuperar entidades de datos relacionadas. Un enlace de datos dado 237 puede incluir consultas para ejecutar contra una fuente de datos del grupo 160 mediante el uso de la semilla como parámetros de entrada. Por ejemplo, si la semilla 212-1 fuera un número de cuenta, entonces el enlace de datos 237 podría recuperar los datos que identifican al propietario de la cuenta con el número de cuenta. Después de recuperar esta información, el motor de agrupamiento 120 agregaría la entidad de datos del cliente al grupo como una entidad de datos relacionada y el atributo del propietario de la cuenta como el enlace entre la semilla 212-1 y

la entidad de datos relacionada. Después de recuperar las entidades de datos relacionadas, el motor de agrupamiento 120 las agrega al grupo 252.

En la etapa 620, el motor de agrupamiento 120 determina si la estrategia de agrupamiento 232-2 se ejecuta completamente. Si es así, el método 600 vuelve a la etapa 615 para ejecutar enlaces de datos adicionales para una semilla dada. Una vez que se ejecuta la estrategia de agrupamiento para esa semilla, el motor de agrupamiento 120 puede determinar y asignar una puntuación a ese grupo (en relación con una estrategia de puntuación especificada). Después de generar grupos para un grupo de semillas, dichos grupos pueden ordenarse o clasificarse con base en los puntajes relativos. Hacerlo permite que un analista identifique y evalúe rápidamente los grupos determinados para representar un alto riesgo de fraude (o tener grandes cantidades en riesgo).

En la etapa 625, el motor de agrupamiento 120 almacena el grupo 252 en la lista de grupos 250. La lista de grupos 250 es una colección de tablas dentro de una base de datos relacional, donde una tabla puede incluir la semilla y las entidades de datos relacionadas del grupo 252 y otra tabla puede incluir enlaces entre las entidades de datos relacionadas del grupo 252. En la etapa 630, el motor de agrupamiento 120 determina si hay más semillas 212 para analizar en la lista de semillas 210. Si es así, el método 600 vuelve a la etapa 605 para generar otro grupo a partir de la siguiente semilla. De lo contrario, el método 600 termina. Tenga en cuenta que, aunque el método 600 describe la generación de un solo grupo, un experto en la técnica reconocerá que el proceso de generación de grupos ilustrado por el método 600 puede realizarse en paralelo.

La Figura 7 es un diagrama de flujo de las etapas del método para puntuar grupos, de acuerdo con una modalidad. Aunque las etapas del método se describen junto con los sistemas de las Figuras 1 y 4, los expertos en la técnica entenderán que cualquier sistema configurado para realizar las etapas del método, en cualquier orden, está dentro del alcance de la presente invención.

Como se muestra, el método 700 comienza en la etapa 705, donde el motor de agrupamiento 120 recupera una estrategia de puntuación 442 y un grupo 252 (por ejemplo, un grupo recién creado mediante el uso del método 600 de la Figura 6). En otros casos, el motor de agrupamiento 120 puede recuperar la estrategia de puntuación 442 asociada con un grupo almacenado. Otras alternativas incluyen un analista que selecciona una estrategia de puntuación 442 a través del cliente 135, el motor de agrupamiento 120 a través de la IU del análisis de agrupamiento 500, una secuencia de comandos o un archivo de configuración. El motor de agrupamiento 120 recupera la estrategia de puntuación seleccionada 442 del almacén de estrategia de puntuación 440. El motor de agrupamiento 120 recupera el grupo 252 de la lista de grupos 250.

En la etapa 710, el motor de agrupamiento 120 ejecuta la estrategia de puntuación 442 contra el grupo 252. La estrategia de puntuación 442 especifica las características de las entidades de datos relacionadas dentro del grupo 252 a agregar. El motor de agrupamiento 120 ejecuta la estrategia de puntuación 442 agregando las características especificadas juntas para determinar una puntuación. Por ejemplo, el motor de agrupamiento 120 podría agregar los saldos de cuenta de entidades de datos relacionadas que son entidades de datos de cuenta. En tal caso, la cantidad total de dólares incluidos en los saldos de las entidades de datos de cuenta del grupo 252 podría ser la puntuación del grupo 252.

En la etapa 715, el motor de agrupamiento 120 almacena la puntuación con el grupo 252 en la lista de grupos 250. En la etapa 720, el motor de agrupamiento 120 determina si hay más grupos 252 para puntuar. Por ejemplo, en una modalidad, se puede volver a puntuar un conjunto de grupos mediante el uso de una estrategia de puntuación actualizada. En otros casos, el motor de agrupamiento puede puntuar cada grupo cuando se crea a partir de una semilla (en función de una generación de grupo dada y la estrategia de puntuación correspondiente). Si quedan más grupos para puntuar (o volver a puntuar), el método 700 vuelve a la etapa 705.

En la etapa 725, el motor de agrupamiento 125 clasifica los grupos 252 de acuerdo con las puntuaciones de los grupos 252. Por ejemplo, después de volver a puntuar un conjunto de grupos (o después de puntuar un conjunto de grupos generados a partir de un conjunto de semillas), el motor de agrupamiento 125 puede clasificar los grupos 252 de la puntuación más alta a la más baja. La calificación se puede usar para ordenar una presentación de resúmenes de los grupos 252 presentados al analista. El analista puede confiar en la calificación y las puntuaciones para determinar qué grupos 252 analizar primero. La calificación y la clasificación generalmente se pueden realizar a pedido cuando un analista busca un grupo para investigar. Por lo tanto, la calificación no necesita suceder al mismo tiempo que la puntuación. Y además, los grupos pueden puntuarse (y luego clasificarse) mediante el uso de diferentes estrategias de clasificación.

La Figura 8 ilustra componentes de un sistema informático de servidor 110, de acuerdo con una modalidad. Como se muestra, el sistema informático de servidor 110 incluye una unidad central de procesamiento (CPU) 860, una interfaz de red 850, una memoria 820 y un almacenamiento 830, cada uno conectado a una interconexión (bus) 840. El sistema informático de servidor 110 puede incluir además una interfaz de dispositivos de E/S 870 para conectar dispositivos de E/S 875 (por ejemplo, teclado, pantalla y ratón) al sistema informático 110. Además, en el contexto de esta descripción, los elementos informáticos mostrados en el sistema informático de servidor 110 pueden corresponder a un sistema informático físico (por ejemplo, un sistema en un centro de datos) o pueden ser una instancia informática virtual que se ejecuta dentro de una nube informática.

La CPU 860 recupera y ejecuta instrucciones de programación almacenadas en la memoria 820, así como también almacena y recupera datos de aplicaciones que residen en la memoria 820. El bus 840 se usa para transmitir las

instrucciones de programación y los datos de las aplicaciones entre la CPU 860, la interfaz de dispositivos de E/S 870, el almacenamiento 830, una interfaz de red 850 y la memoria 820. Tenga en cuenta que la CPU 860 se incluye como representación de una sola CPU, múltiples CPU, una sola CPU que tiene múltiples núcleos de procesamiento, una CPU con una unidad de administración de memoria asociada, y similares. Y la memoria 820 se incluye generalmente para representación de una memoria de acceso aleatorio. El almacenamiento 830 puede ser un dispositivo de almacenamiento de unidad de disco. Aunque se muestra como una unidad sencilla, el almacenamiento 830 puede ser una combinación de dispositivos de almacenamiento fijos y/o removibles, tales como unidades de discos, tarjetas de memoria extraíbles, o almacenamiento óptico, almacenamiento conectado a la red (NAS), o un almacenamiento de red de área (SAN).

Ilustrativamente, la memoria 820 incluye una lista de semillas 210, un motor de agrupamiento 120, una lista de grupos 250 y un motor de flujo de trabajo 125. El motor de agrupamiento 120 incluye una estrategia de agrupamiento 232-2. La estrategia de agrupamiento particular 232-2 incluye enlaces de datos 237-1, 237-2 y 237-3, con los cuales el motor de agrupamiento 120 accede a la fuente de datos del grupo 160. El motor de flujo de trabajo 125 incluye una estrategia de puntuación 442-1.

Ilustrativamente, el almacenamiento 830 incluye un almacén de estrategia de agrupamiento 230, un almacén de enlaces de datos 835 y un almacén de estrategia de puntuación 440. Como se discutió, el almacén de estrategias de agrupamiento 230 puede incluir una colección de diferentes estrategias de agrupamiento 232, tales como la estrategia de agrupamiento 232-2. El almacén de estrategias de agrupamiento 230 puede ser un directorio que incluye las estrategias de agrupamiento 232-1, 232-2 ... 232-N como módulos distintos. El almacén de estrategias de puntuación 440 puede incluir una colección de diferentes estrategias de puntuación 442, tales como la estrategia de puntuación 442-2 y también puede ser un directorio de módulos distintos. El almacén de enlace de datos 835 incluye enlaces de datos 237-1, 237-2 ... 237-M, que también puede almacenarse como módulos distintos dentro de un directorio.

Aunque se muestra en la memoria 820, la lista de semillas 210, el motor de agrupamiento 120, la lista de grupos 250 y el motor de flujo de trabajo 125, pueden almacenarse en la memoria 820, almacenamiento 830 o dividirse entre la memoria 820 y el almacenamiento 830. Del mismo modo, las copias de la estrategia de agrupamiento 232-2, el enlace de datos 237-1, 237-2 y 237-3 y la estrategia de puntuación 442-2 pueden almacenarse en la memoria 820, el almacenamiento 830 o dividirse entre la memoria 820 y el almacenamiento 830.

Tenga en cuenta que, si bien el fraude financiero con cuentas de tarjetas de crédito se utiliza como un ejemplo de referencia principal en la discusión anterior, un experto en la técnica reconocerá que las técnicas descritas en la presente descripción pueden adaptarse para su uso con una variedad de conjuntos de datos. Por ejemplo, la información de los registros de datos de los sistemas en línea podría evaluarse como semillas para mejorar la seguridad cibernética. En tal caso, una semilla podría ser una dirección IP sospechosa, una cuenta de usuario comprometida, etc. Desde las semillas, los datos de registro, los registros de DHCP, las capturas de paquetes de listas negras de IP, los registros de aplicaciones web y otros registros de servidores y bases de datos podrían usarse para crear grupos de actividad relacionados con las semillas de sospechas. Otros ejemplos incluyen el análisis de calidad de datos usado para agrupar transacciones procesadas a través de un sistema informático (ya sea financiero o de otro tipo).

Se han descrito modalidades de la presente descripción que se refieren a la generación automática de estructuras de datos agrupados eficientes en memoria y, más específicamente, a la selección automática de una entidad de datos inicial de interés, agregando la entidad de datos inicial a la estructura de datos agrupados eficiente en memoria y determinar y agregar una o más entidades de datos relacionadas al grupo. Como se describió anteriormente, en diversas modalidades, un grupo generado puede incluir muchas menos entidades de datos en comparación con una gran colección de elementos de datos que pueden o no estar relacionados entre sí. Esto puede deberse a que, por ejemplo, las entidades de datos incluidas en un grupo solo pueden incluir aquellas entidades de datos que están relacionadas entre sí y que pueden ser relevantes para una investigación en particular. Por consiguiente, en diversas modalidades, el procesamiento de grupos generados puede ser altamente eficiente porque, por ejemplo, una investigación de fraude dada por un analista puede requerir solo almacenamiento en la memoria de una estructura de datos de agrupamiento único. Además, varias entidades de datos en un grupo pueden ser varios órdenes de magnitud más pequeñas que en la gran colección de elementos de datos que pueden o no estar relacionados entre sí porque solo las entidades de datos relacionadas entre sí están incluidas en el grupo.

Aunque lo anterior se dirige a las modalidades de la presente invención, otras modalidades adicionales de la invención se pueden concebir sin apartarse del alcance básico de las mismas, y el alcance de las mismas se determina por las reivindicaciones siguientes. Por ejemplo, los aspectos de la presente invención pueden implementarse en soporte físico o programa informático o en una combinación de soporte físico y programa informático. Una modalidad de la invención puede implementarse como un producto de programa para usar con un sistema informático. El (los) programa(s) del producto del programa define(n) las funciones de las modalidades (incluidos los métodos descritos en la presente descripción) y puede estar contenido en una variedad de medios de almacenamiento legibles por ordenador. Los medios de almacenamiento ilustrativos legibles por ordenador incluyen, entre otros: (i) medios de almacenamiento no grabables (por ejemplo, dispositivos de memoria de solo lectura dentro de un ordenador, como discos CD-ROM legibles por una unidad de CD-ROM, memoria flash, chips ROM o cualquier tipo de memoria semiconductora no volátil de estado sólido) en la que la información se almacena permanentemente; y (ii) medios de almacenamiento grabables (por ejemplo, unidad

de disco duro o cualquier tipo de memoria semiconductora de acceso aleatorio de estado sólido) en la que se almacena información alterable.

Es particularmente ventajoso si la semilla se selecciona automáticamente, aunque en otras modalidades la semilla puede ser seleccionada por un analista u otro usuario y la semilla luego usada de la misma manera,

La invención se ha descrito anteriormente de acuerdo con las modalidades específicas. Sin embargo, personas no expertas en la técnica entenderán que se pueden hacer varias modificaciones y cambios a los mismos sin apartarse del alcance más amplio de la invención como se establece en las reivindicaciones adjuntas. En consecuencia, la descripción y los dibujos anteriores deben considerarse con un sentido ilustrativo y no restrictivo. En lugar de eso, el alcance de la presente invención se define por las siguientes reivindicaciones.

REIVINDICACIONES

1. Un método implementado por ordenador para generar un grupo de entidades de datos relacionadas, el método comprende:
establecer comunicación con uno o más almacenes de datos electrónicos (820, 830) almacenando una pluralidad de entidades de datos y atributos de entidad de datos respectivos, uno o más almacenes de datos electrónicos en comunicación con uno o más procesadores informáticos de soporte físco (860), uno o más procesadores informáticos de soporte físico configurados con instrucciones ejecutables por ordenador específicas;
generar, para cada una de las entidades de datos semilla seleccionadas de la pluralidad de entidades de datos, y por uno o más procesadores de informáticos de soporte físico, un grupo de entidades de datos al menos por:
designar automáticamente la entidad de datos semilla (302) como una entidad de datos inicial del grupo de entidades de datos (252, 252-1);
acceder, basado en una estrategia de agrupamiento (232), a dos o más protocolos de búsqueda (237-1, 237-2, ..., 237-M);
ejecutar un primero de los dos o más protocolos de búsqueda en uno o más almacenes de datos electrónicos para identificar una o más entidades de datos relacionadas con la entidad de datos semilla, en donde la ejecución del primero de los dos o más protocolos de búsqueda comprende:
identificar, por uno o más procesadores informáticos de soporte físico, al menos un atributo de entidad de datos (303-1) asociado con la entidad de datos semilla;
evaluar, por uno o más procesadores informáticos de soporte físco, la pluralidad de entidades de datos para determinar una o más entidades de datos (305-1) que comparten, al menos, un atributo de entidad de datos con la entidad de datos semilla; y
agregar una o más entidades de datos al grupo de entidades de datos; y
ejecutar iterativamente cada uno de los otros protocolos de búsqueda de los dos o más protocolos de búsqueda en uno o más almacenes de datos electrónicos para identificar una o más entidades de datos adicionales (305-2, ..., 305-6) relacionadas con una o más entidades de datos previamente agregadas al grupo de entidades de datos, en donde la ejecución de cada uno de los otros protocolos de búsqueda de los dos o más protocolos de búsqueda comprende:
identificar, por uno o más procesadores informáticos de soporte físco, al menos un atributo de entidad de datos (303-2, ..., 303-6) asociado con al menos una de las una o más entidades de datos agregadas previamente al grupo de entidades de datos, en donde el protocolo de búsqueda usa el atributo de entidad de datos identificado al menos como un parámetro de entrada de datos para el protocolo de búsqueda;
evaluar, por uno o más procesadores informáticos de soporte físco y basándose en la ejecución del protocolo de búsqueda, de la pluralidad de entidades de datos para determinar una o más entidades de datos adicionales que comparten el al menos un atributo de entidad de datos con al menos uno de uno o más entidades de datos previamente agregadas al grupo de entidades de datos; y
agregar una o más entidades de datos adicionales al grupo de entidades de datos (252-1); y
almacenar grupo de entidad de datos.

2. El método de la reivindicación 1 que comprende, además:
comparar, mediante uno o más procesadores informáticos de soporte físico, entidades de datos asociadas con el grupo de entidades de datos (252-1) con entidades de datos asociadas con un segundo grupo de entidades de datos (252-C); y
en respuesta a la determinación de que al menos una entidad de datos asociada con el grupo de entidades de datos comparte un atributo y/o está relacionado con al menos una entidad de datos asociada con el segundo grupo de entidades de datos, fusionar el grupo de entidades de datos (252-1) y el segundo grupo de entidades de datos (252-C).

3. El método de la reivindicación 1, en donde el primer protocolo de búsqueda usa la semilla de entidad de datos como una entrada de parámetro de datos al primer protocolo de búsqueda.

4. El método de la reivindicación 3, en donde el primer protocolo de búsqueda busca entidades de datos particulares en un primer almacén de datos electrónicos y los otros protocolos de búsqueda de los dos o más protocolos de búsqueda buscan entidades de datos particulares en un segundo almacén de datos electrónicos.

5. El método de cualquiera de las reivindicaciones 3 a 4, que comprende, además: asignar una puntuación de clasificación al grupo de entidades de datos, en donde la puntuación de clasificación se usa para ordenar el grupo de entidades de datos con relación a una pluralidad de otros grupos de entidades de datos generados a partir de semillas de entidades de datos respectivas de acuerdo con la estrategia del grupo.

6. Un programa informático que comprende las instrucciones legibles por máquina que cuando se ejecutan por un aparato informático que hace que realice el método de cualquiera de las reivindicaciones anteriores.

7. Un aparato que comprende:
uno o más almacenes de datos electrónicos (820, 830) que almacenan una pluralidad de entidades de datos y atributos de entidad de datos respectivos; y

uno o más procesadores informáticos de soporte físico (860) en comunicación con uno o más almacenes de datos electrónicos y configurados con instrucciones ejecutables específicas del ordenador, las instrucciones configuradas para hacer que uno o más procesadores informáticos de soporte físico:
genere, para cada una de las entidades de datos semilla seleccionadas de la pluralidad de entidades de datos, un grupo de entidades de datos al menos por:
designar automáticamente la entidad de datos semilla (302) como una entidad de datos inicial del grupo de entidades de datos (252, 252-1);
acceder, basado en una estrategia de agrupamiento (232), a dos o más protocolos de búsqueda (237-1, 237-2, ..., 237-M);
ejecutar un primero de los dos o más protocolos de búsqueda en uno o más almacenes de datos electrónicos para identificar una o más entidades de datos relacionadas con la entidad de datos semilla, en donde la ejecución del primero de los dos o más protocolos de búsqueda comprende:
identificar al menos un atributo de entidad de datos (303-1) asociado con la entidad de datos semilla;
evaluar la pluralidad de entidades de datos para determinar una o más entidades de datos (305-1) que comparten al menos un atributo de entidad de datos con la entidad de datos semilla; y
agregar una o más entidades de datos al grupo de entidades de datos; y
ejecutar iterativamente cada uno de los otros protocolos de búsqueda de los dos o más protocolos de búsqueda en uno o más almacenes de datos electrónicos para identificar una o más entidades de datos adicionales (305-2, ..., 305-6) relacionadas con una o más entidades de datos previamente agregadas al grupo de entidades de datos, en donde la ejecución de cada uno de los otros protocolos de búsqueda de los dos o más protocolos de búsqueda comprende:
identificar al menos un atributo de entidad de datos (303-2, ..., 303-6) asociado con al menos una de una o más entidades de datos agregadas previamente al grupo de entidades de datos, en donde el protocolo de búsqueda usa al menos la entidad de datos identificada atribuida como entrada de parámetro de datos al protocolo de búsqueda;
evaluar, basándose en la ejecución del protocolo de búsqueda, la pluralidad de entidades de datos para determinar una o más entidades de datos adicionales que comparten al menos un atributo de entidad de datos con al menos una de una o más entidades de datos agregadas previamente al grupo de entidades de datos; y
agregar una o más entidades de datos adicionales al grupo de entidades de datos (252-1); y
almacenar el grupo de entidades de datos en uno o más almacenes de datos electrónicos.

8. El aparato de la reivindicación 7, las instrucciones configuradas además para hacer que uno o más procesadores informáticos de soporte físico:
compare las entidades de datos asociadas con el grupo de entidades de datos (252-1) con las entidades de datos asociadas con un segundo grupo de entidades de datos (252-C); y
en respuesta a la determinación de que al menos una entidad de datos asociada con el grupo de entidades de datos comparte un atributo y/o está relacionado con al menos una entidad de datos asociada con el segundo grupo de entidades de datos, combine el grupo de entidades de datos (252-1) y el segundo grupo de entidades de datos (252-C).

9. El aparato de la reivindicación 7, en donde el primer protocolo de búsqueda usa la semilla de entidad de datos como una entrada de parámetro de datos al primer protocolo de búsqueda.

10. El aparato de la reivindicación 9, en donde el primer protocolo de búsqueda busca entidades de datos particulares en un primer almacén de datos electrónicos y los otros protocolos de búsqueda de los dos o más protocolos de búsqueda buscan entidades de datos particulares en un segundo almacén de datos electrónicos.

11. El aparato de cualquiera de las reivindicaciones de la 9 a la 10, las instrucciones configuradas además para hacer que uno o más procesadores informáticos de soporte físico:
asigne una puntuación de clasificación al grupo de entidades de datos, en donde la puntuación de clasificación se usa para ordenar el grupo de entidades de datos con relación a una pluralidad de otros grupos de entidades de datos generados a partir de semillas de entidades de datos respectivas de acuerdo con la estrategia de agrupamiento.

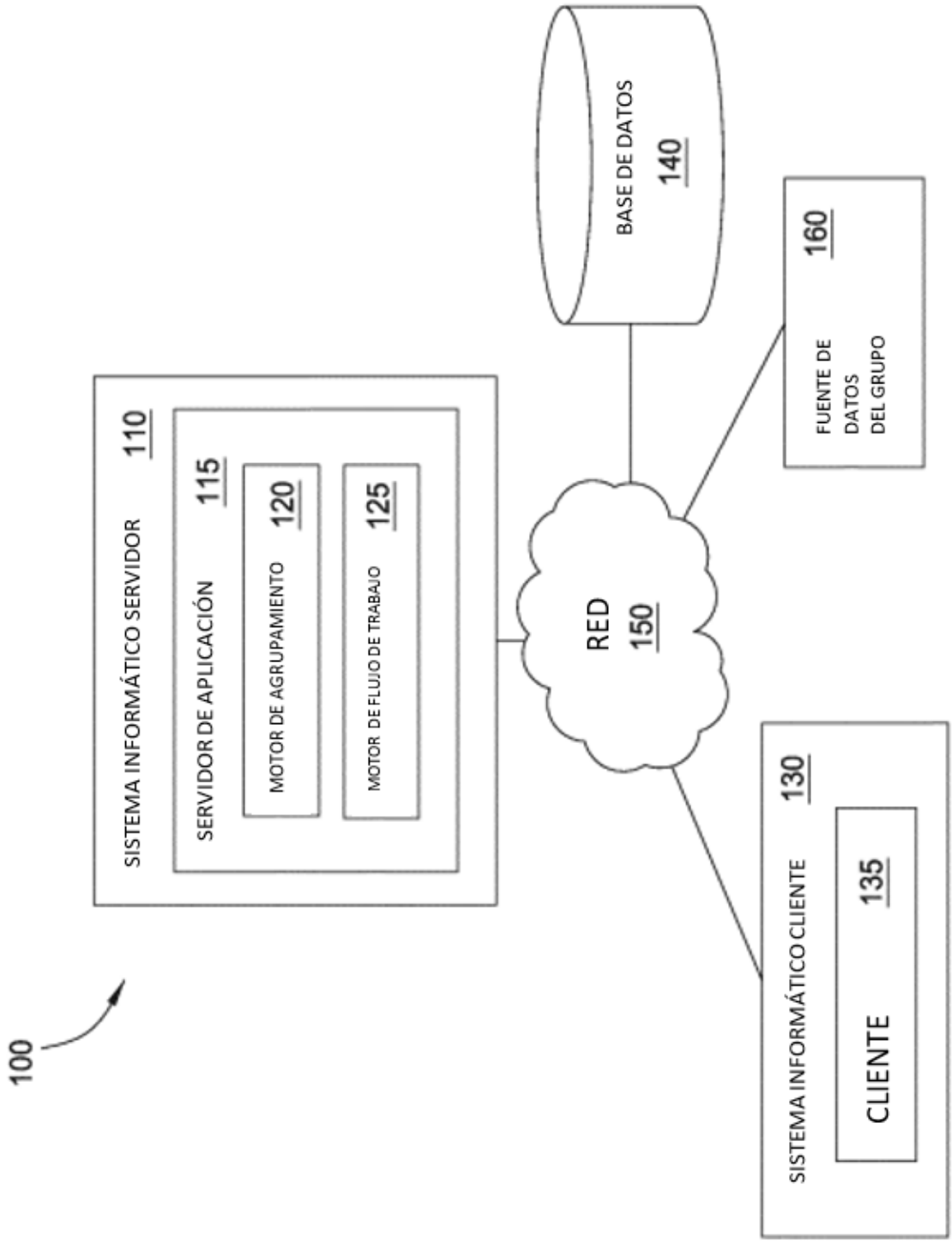
100
SISTEMA INFORMÁTICO SERVIDOR 110
SERVIDOR DE APLICACIÓN 115
MOTOR DE AGRUPAMIENTO 120
MOTOR DE FLUJO DE TRABAJO 125
RED 150
BASE DE DATOS 140
FUENTE DE DATOS DEL GRUPO 160
SISTEMA INFORMÁTICO CLIENTE 130
CLIENTE 135

FIGURA 1

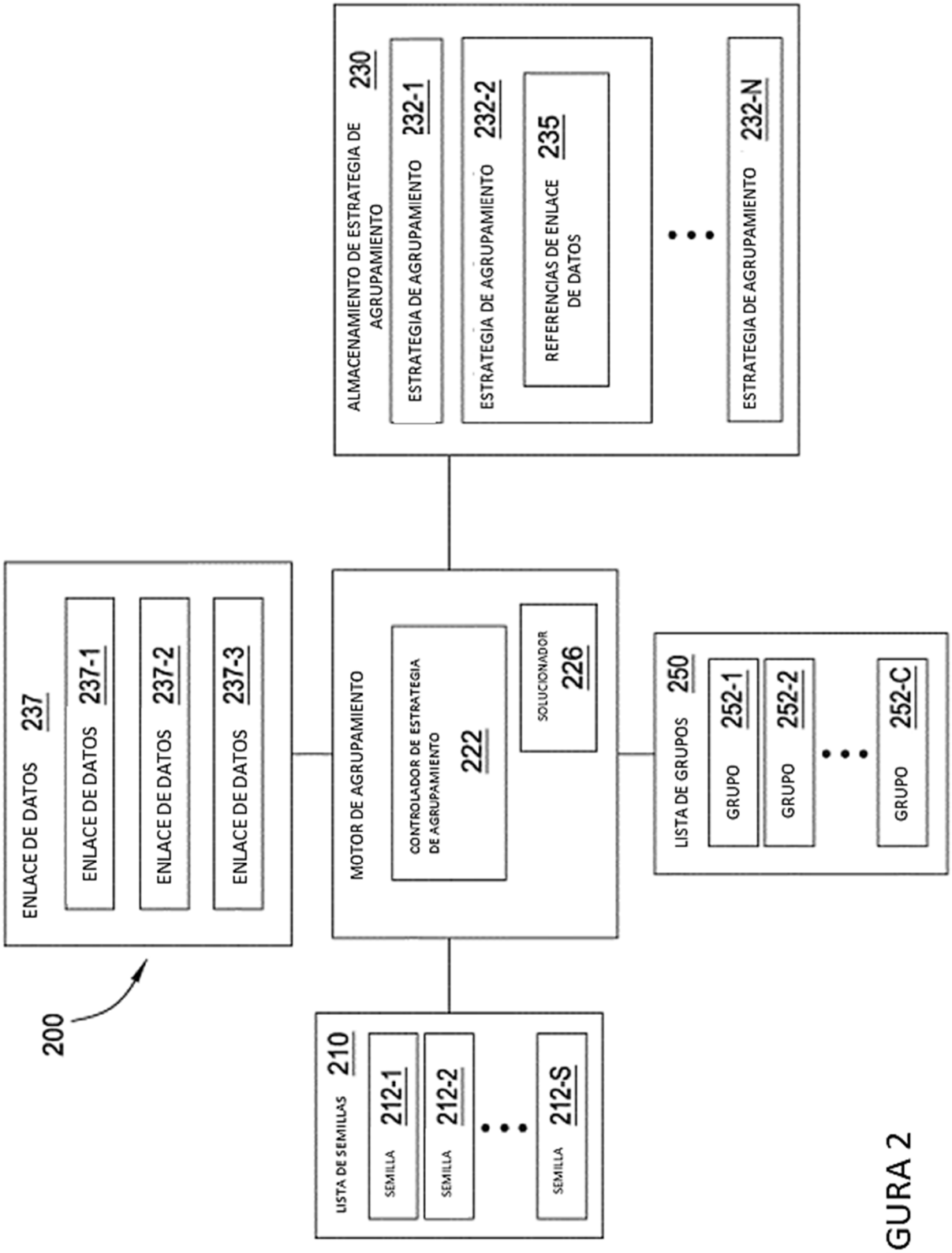
200
ENLACE DE DATOS 237
ENLACE DE DATOS 237-1
ENLACE DE DATOS 237-2
ENLACE DE DATOS 237-3
LISTA DE SEMILLAS 210
SEMILLA 212-1
SEMILLA 212-2
SEMILLA 212-S
MOTOR DE AGRUPAMIENTO
CONTROLADOR DE ESTRATEGIA DE AGRUPAMIENTO 222
SOLUCIONADOR 226
ALMACENAMIENTO DE ESTRATEGIA DE AGRUPAMIENTO 230
ESTRATEGIA DE AGRUPAMIENTO 232-1
ESTRATEGIA DE AGRUPAMIENTO 232-2
REFERENCIAS DE ENLACE DE DATOS 235
ESTRATEGIA DE AGRUPAMIENTO 232-N
LISTA DE GRUPOS 250
GRUPO 252-1
GRUPO 252-2
GRUPO 252-C

FIGURA 2

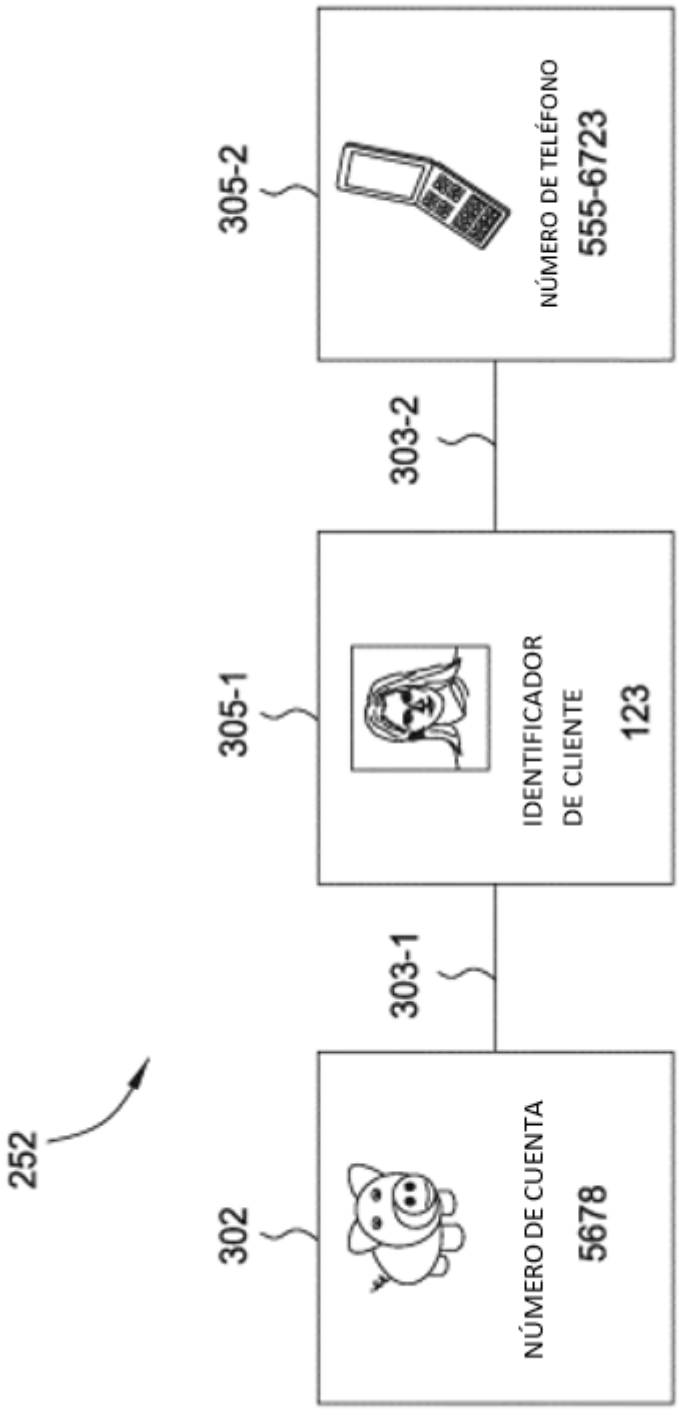
252
302
NÚMERO DE CUENTA
5678
303-1
305-1
IDENTIFICADOR
DE CLIENTE
123
303-2
305-2
NÚMERO DE TELÉFONO
555-6723

FIGURA 3A

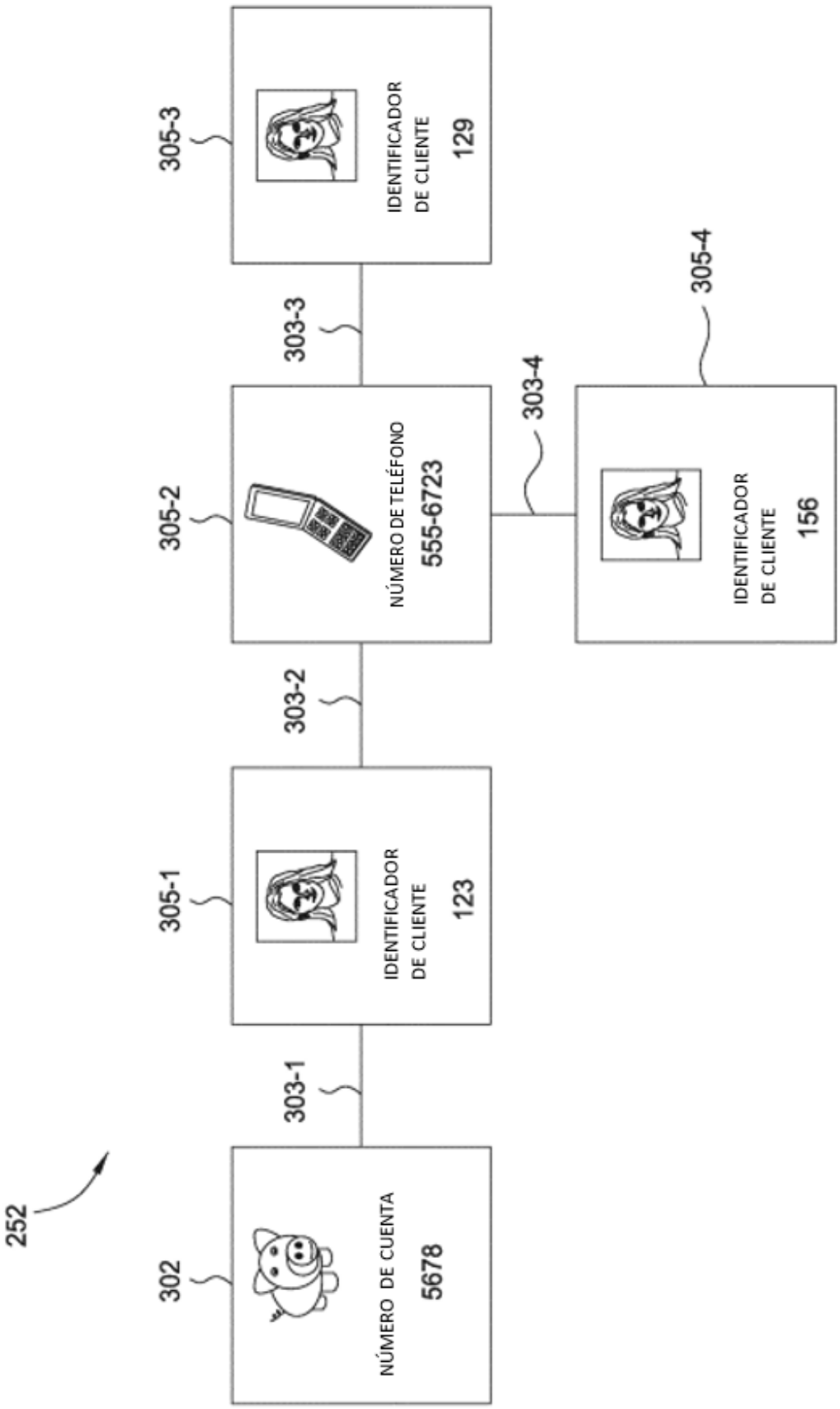
252
302
NÚMERO DE CUENTA
5678
303-1
305-1
IDENTIFICADOR
DE CLIENTE
123
303-2
305-2
NÚMERO DE TELÉFONO
555-6723
303-3
305-3
IDENTIFICADOR
DE CLIENTE
129
303-4
305-4
IDENTIFICADOR
DE CLIENTE
156

FIGURA 3B

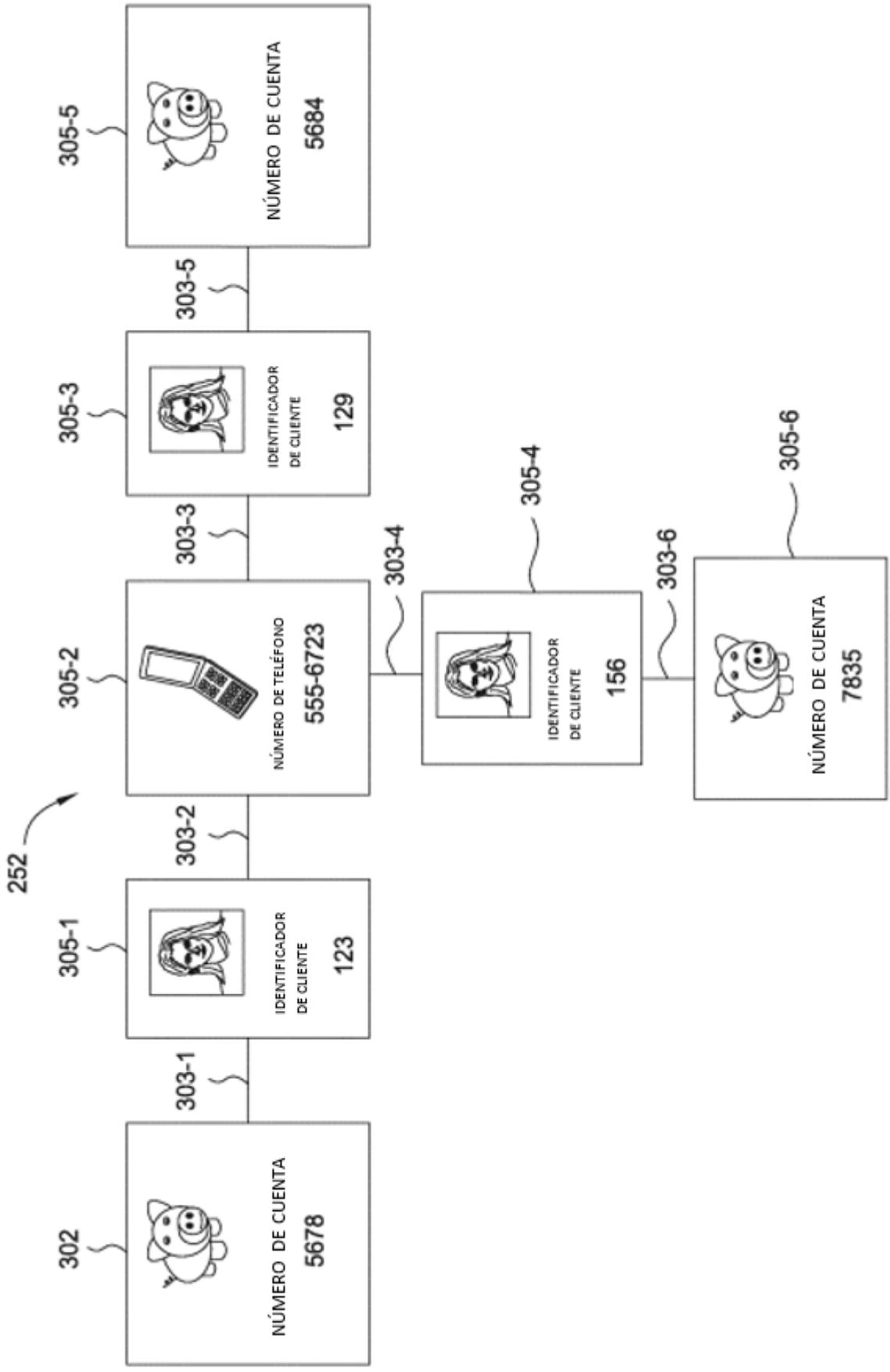
252
302
NÚMERO DE CUENTA
5678
303-1
305-1
IDENTIFICADOR
DE CLIENTE
123
303-2
305-2
NÚMERO DE TELÉFONO
555-6723
303-3
305-3
IDENTIFICADOR
DE CLIENTE
129
303-5
305-5
NÚMERO DE CUENTA
5684
303-4
305-4
IDENTIFICADOR
DE CLIENTE
156
303-6
305-6
NÚMERO DE CUENTA
7835

FIGURA 3C

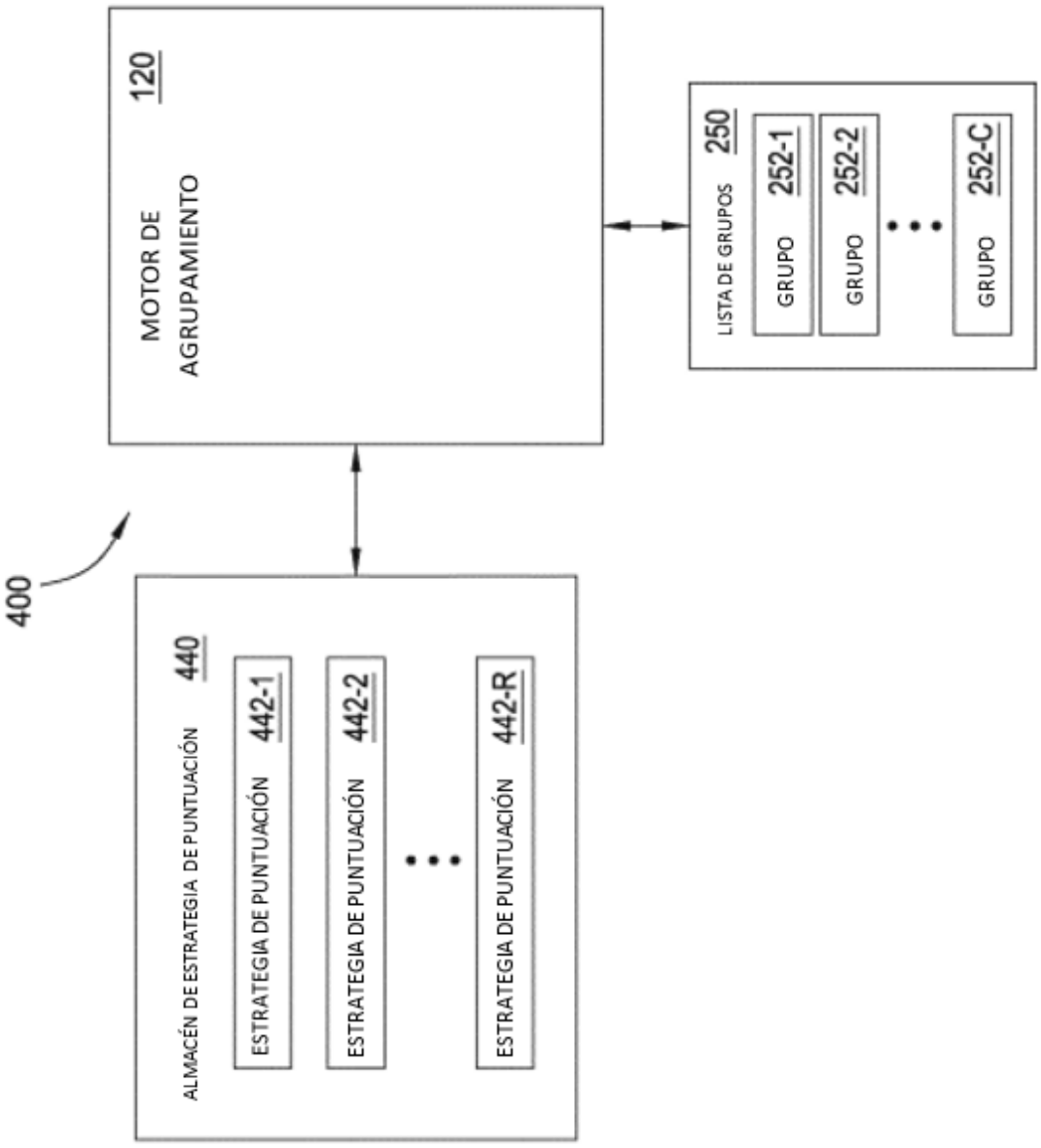
400
MOTOR DE AGRUPAMIENTO 120
LISTA DE GRUPOS 250
GRUPO 252-1
GRUPO 252-2
GRUPO 252-C
ALMACÉN DE ESTRATEGIA DE PUNTUACIÓN 440
ESTRATEGIA DE PUNTUACIÓN 442-1
ESTRATEGIA DE PUNTUACIÓN 442-2
ESTRATEGIA DE PUNTUACIÓN 442-R

FIGURA 4

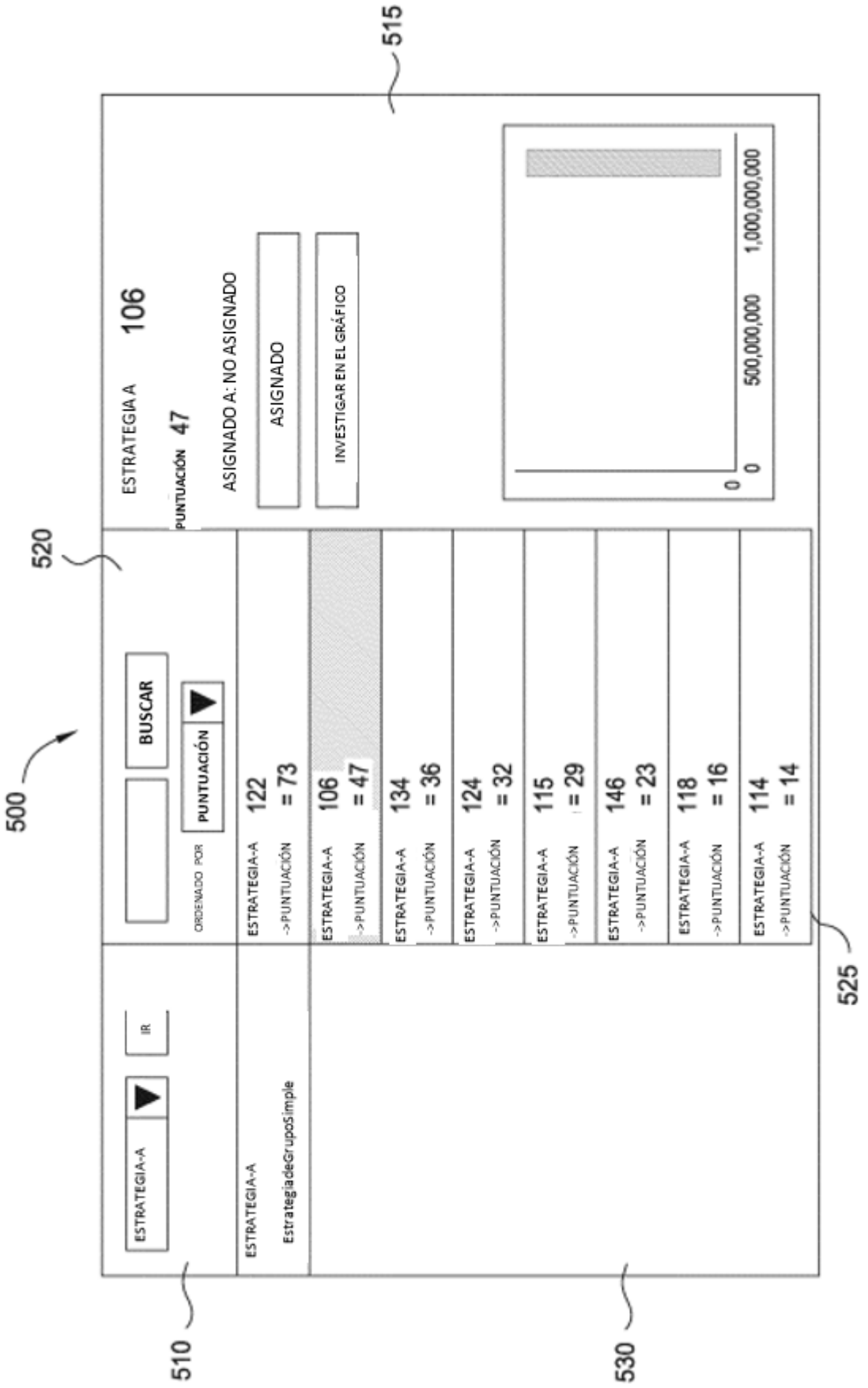
500
510
520
530
515
525
ESTRATEGIA-A
IR
ESTRATEGIA-A
EstrategiadeGruposSimple
ORDENADO POR
PUNTUACIÓN
BUSCAR
ESTRATEGIA-A
->PUNTUACIÓN
122
= 73
106
= 47
134
= 36
124
= 32
115
= 29
146
= 23
118
= 16
114
= 14
ESTRATEGIA A
106
PUNTUACIÓN
47
ASIGNADO A: NO ASIGNADO
ASIGNADO
INVESTIGAR EN EL GRÁFICO
0
0
500,000,000
1,000,000,000

FIGURA 5

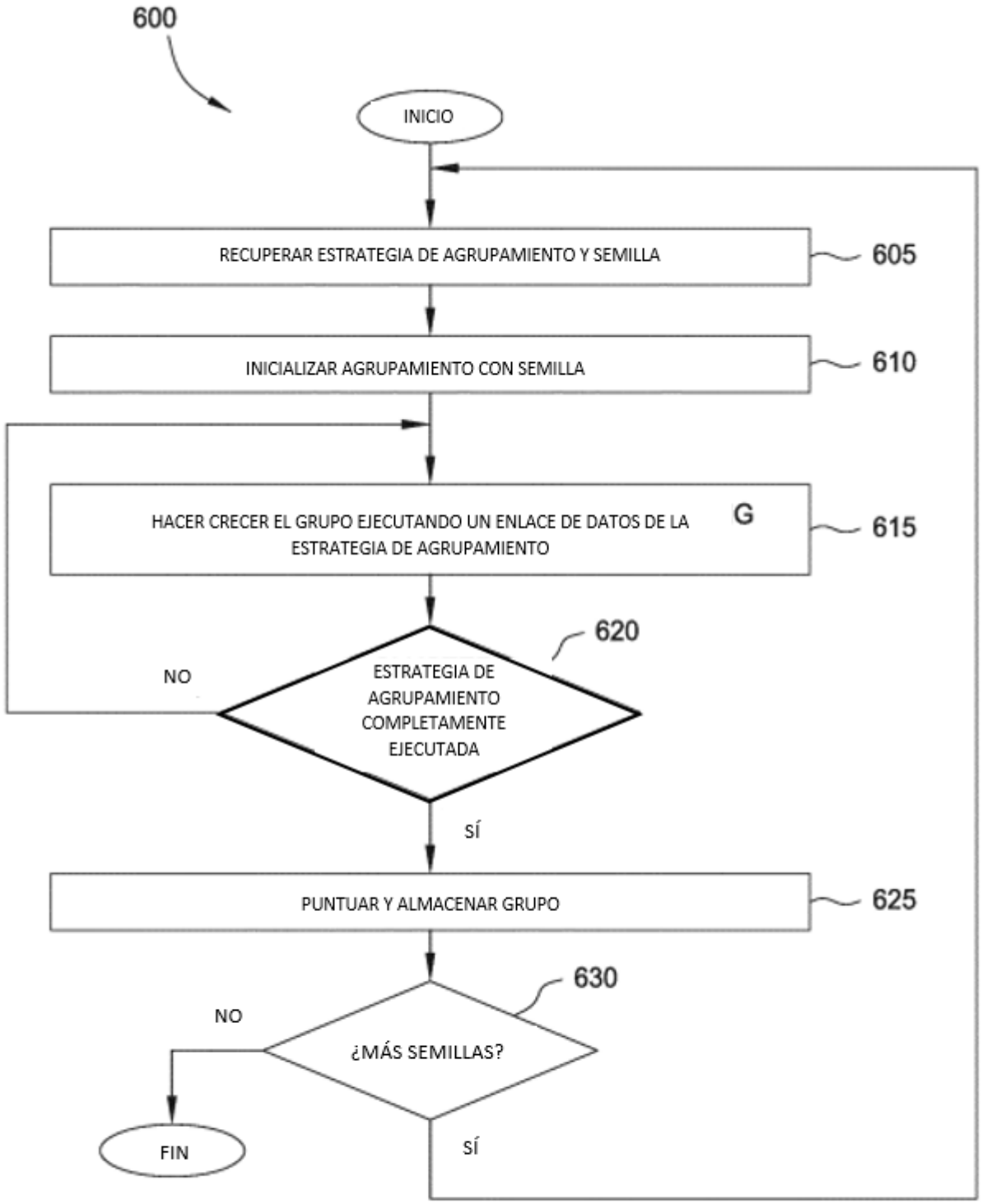
600
INICIO
RECUPERAR ESTRATEGIA DE AGRUPAMIENTO Y SEMILLA
605
INICIALIZAR AGRUPAMIENTO CON SEMILLA
610
HACER CRECER EL GRUPO EJECUTANDO UN ENLACE DE DATOS DE LA ESTRATEGIA DE AGRUPAMIENTO
G
615
620
ESTRATEGIA DE AGRUPAMIENTO COMPLETAMENTE EJECUTADA
NO
SÍ
PUNTUAR Y ALMACENAR GRUPO
625
630
¿MÁS SEMILLAS?
NO
SÍ
FIN

FIGURA 6

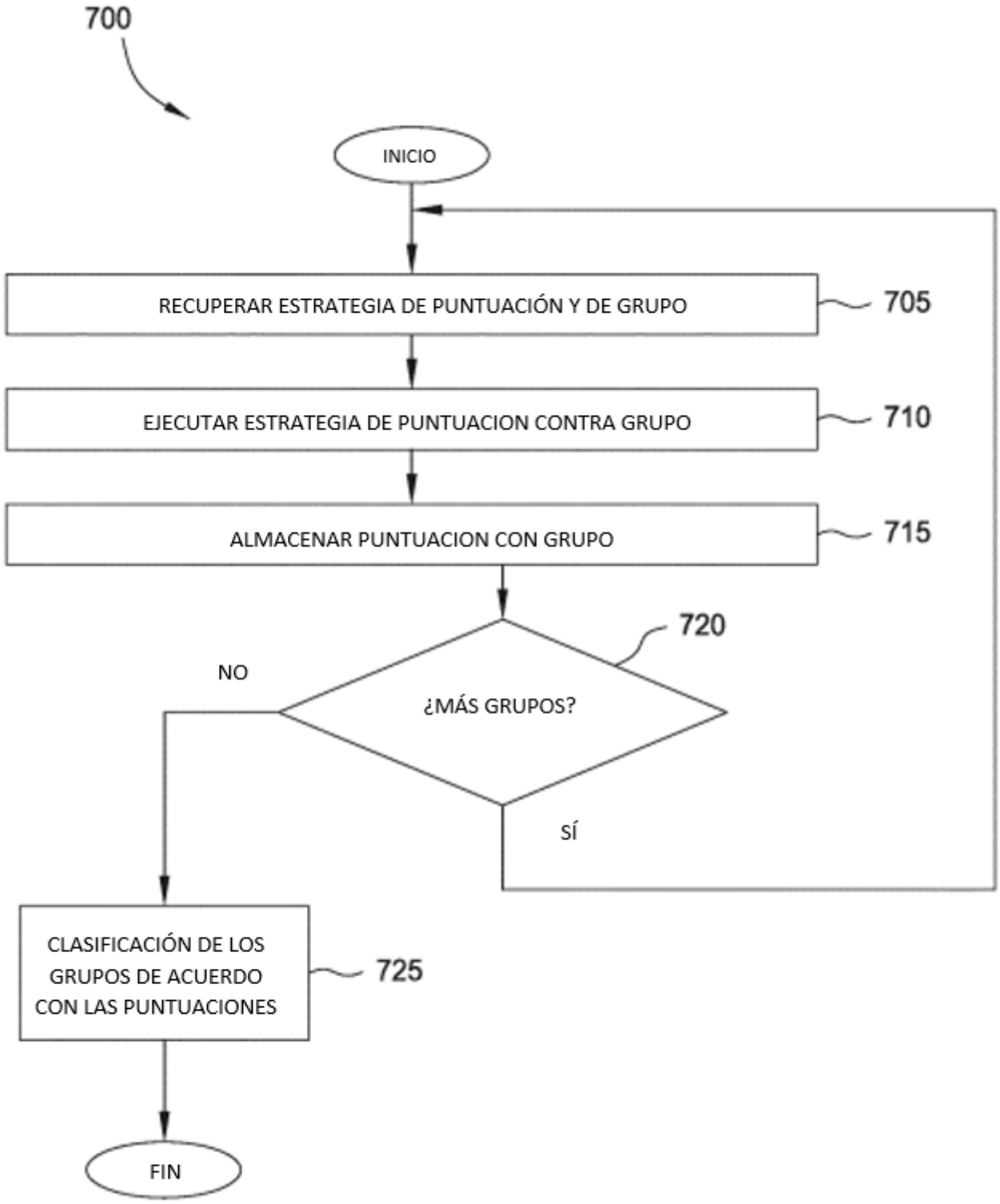
700
INICIO
RECUPERAR ESTRATEGIA DE PUNTUACIÓN Y DE GRUPO
705
EJECUTAR ESTRATEGIA DE PUNTUACION CONTRA GRUPO
710
ALMACENAR PUNTUACION CON GRUPO
715
720
¿MÁS GRUPOS?
NO
SÍ
CLASIFICACIÓN DE LOS GRUPOS DE ACUERDO CON LAS PUNTUACIONES
725
FIN

FIGURA 7

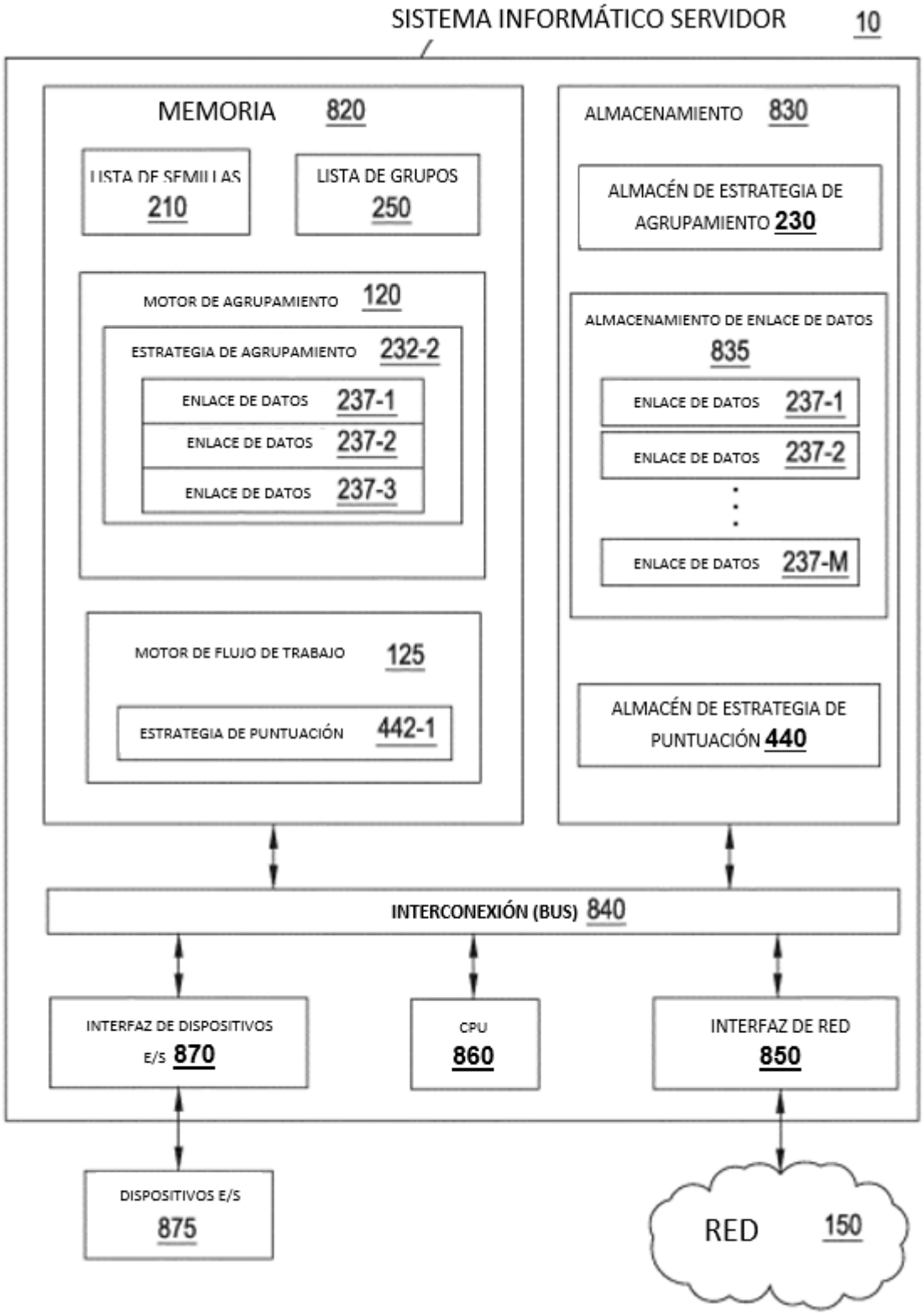
SISTEMA INFORMÁTICO SERVIDOR 10
MEMORIA 820
LISTA DE SEMILLAS 210
LISTA DE GRUPOS 250
MOTOR DE AGRUPAMIENTO 120
ESTRATEGIA DE AGRUPAMIENTO 232-2
ENLACE DE DATOS 237-1
ENLACE DE DATOS 237-2
ENLACE DE DATOS 237-3
MOTOR DE FLUJO DE TRABAJO 125
ESTRATEGIA DE PUNTUACIÓN 442-1
ALMACENAMIENTO 830
ALMACÉN DE ESTRATEGIA DE AGRUPAMIENTO 230
ALMACENAMIENTO DE ENLACE DE DATOS 835
ENLACE DE DATOS 237-1
ENLACE DE DATOS 237-2
ENLACE DE DATOS 237-M
ALMACÉN DE ESTRATEGIA DE PUNTUACIÓN 440
INTERCONEXIÓN (BUS) 840
INTERFAZ DE DISPOSITIVOS E/S 870
CPU 860
INTERFAZ DE RED 850
DISPOSITIVOS E/S 875
RED 150

FIGURA 8